

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 1 de 128

**Sistema para la identificación temprana de Varroosis en las abejas implementando
inteligencia artificial.**

Saray Natalia Alvarado Acosta

Martha Lucia Caicedo Morales

Titulación

Trabajo de grado

Director(a):

Julián Rene Muñoz Burbano

Codirector(a):

Corporación Universitaria ComfacaUCA – UnicomfacaUCA

Ingeniería de sistemas

Popayán

Mayo 2023

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 2 de 128

Agradecimientos

Gracias a Dios y a nuestros queridos padres, cuyo amor incondicional y apoyo constante fueron las bases fundamentales para avanzar en nuestro camino académico.

Gracias a nuestra querida universidad UnicomfacaUCA por las oportunidades que nos dio para obtener conocimiento y crecer como profesionales.

Gracias a todos los apicultores que con su invaluable cooperación nos apoyaron durante este valioso proceso. Gracias por enseñarnos y mantener un compromiso con el cuidado de las abejas.

Y por último damos gracias a nosotras mismas, por mantener la perseverancia y firmeza en todo momento.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 3 de 128

Resumen

El presente trabajo de grado tiene como objetivo principal desarrollar un sistema de detección temprana de Varroosis utilizando técnicas de inteligencia artificial como las redes neuronales convolucionales. La Varroosis es una enfermedad parasitaria provocada por el ácaro *Varroa destructor* que debilita el sistema inmunológico de las abejas generando problemas de salud y supervivencia.

El sistema propuesto utiliza un Dataset creado a partir de la recolección de imágenes reales compuestas por abejas sanas y abejas contagiadas de Varroosis. Estas son procesadas por medio de la técnica de transferencia de aprendizaje la cual permite tomar una red pre entrenada del modelo YoloV5, que es una red neuronal Convolucional de detección de objetos en tiempo real. Con esta información es posible clasificar en qué estado de salud se encuentran las abejas. Los resultados de las pruebas demuestran una alta precisión en la detección de abejas sanas o enfermas, demostrando así la eficacia del sistema propuesto. Este proyecto es realizado con el fin de que pueda ser utilizado por apicultores de la región, permitiendo que se logre detectar a tiempo la presencia de la *Varroa* en las colmenas de los apiarios.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 4 de 128

Abstract

The main objective of this thesis is to develop an early detection system for Varroosis using artificial intelligence techniques such as convolutional neural networks. Varroosis is a parasitic disease caused by the *Varroa destructor* mite, which weakens the immune system of bees, leading to health problems and reduced survival.

The proposed system utilizes a dataset created from the collection of real images containing both healthy bees and bees infected with Varroosis. These images are processed using transfer learning techniques, which involve utilizing a pre-trained model YoloV5, a convolutional neural network for real-time object detection. With this information, it becomes possible to classify the health status of the bees.

The test results demonstrate a high accuracy in detecting healthy or sick bees, thus demonstrating the effectiveness of the proposed system. This project is carried out with the purpose of being used by beekeepers in the region, enabling timely detection of Varroa presence in apiary hives.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 5 de 128

Tabla de contenido

Lista de figuras.....	9
Lista de tablas.....	11
Glosario.....	12
Capítulo 1. Introducción.	14
1.1. Determinar la <i>Varroa destructor</i> en colonias de abejas del Macizo Colombiano.....	16
1.2. Propuesta	18
Capítulo 2. Marco Teórico	19
2.1 las Abejas	20
2.1.1. Características.....	21
2.1.1.1. <i>Apis mellifera</i>	22
2.1.2. Importancia.....	22
2.2. La Varroosis en las abejas	23
2.2.2. Historia	23
2.2.3. Morfología.....	24
2.2.4. Ciclo.....	24
2.2.5. Daños	25
2.2.5.1. Daños directos sobre la cría de las abejas.....	26
2.2.5.2. Daños indirectos.	26
2.2.6. Modo de Propagación	26
2.2.7. <i>Varroa destructor</i>	27
2.2.8. Métodos de control más reconocidos	27
2.2.8.1. Tratamientos con productos naturales.....	27
2.2.8.2. Técnicas culturales Colmenas control.	28
2.3. Inteligencia artificial	30
2.4. El Machine Learning.....	30
2.4.1. Tipos de aprendizaje.	31
2.4.1.1. Aprendizaje supervisado.....	31
2.4.1.2. Aprendizaje No supervisado.....	33

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 6 de 128

2.4.2. Modelos de Machine Learning.....	33
2.4.2.1. Modelos lineales.....	33
2.4.2.2. Modelos de árbol.....	33
2.4.2.3. Redes neuronales.....	34
2.4.3. Cuadro comparativo de los modelos de Machine Learning.	35
2.4.4. Redes Neuronales Convolucionales (CNN)	37
2.4.5. Fases de desarrollo Modelos de Machine Learning.	38
2.4.5.1. Fase de entrenamiento.	38
2.4.5.2. Fase de prueba.....	38
2.4.6. Visión Artificial.....	39
2.4.7. Procesamiento digital de imágenes.....	40
2.4.7.1. Pasos para el procesamiento de imágenes.	41
2.4.7.2. Tipos de una imagen.	41
2.4.7.3. Ventajas.	42
2.5. Metodologías Machine learning	42
2.5.1. CRISP-DM (Cross Industry Standard Process for Data Mining)	42
2.5.1.1. Comprensión del problema o negocio.	44
2.5.1.2. Comprensión de datos.....	44
2.5.1.3. Preparación de datos.	44
2.5.1.4. Modelado.	45
2.5.1.5. Evaluación del modelo.	45
2.5.1.6. Implementación del modelo.	46
2.5.2. Proceso de ciencia de datos en equipo (TDSP)	46
2.5.2.1. Principales componentes del TDSP	47
2.5.2.1.1. Una definición de ciclo de vida de la ciencia de datos.....	47
2.5.2.1.2. Una estructura de proyecto estandarizada.....	48
2.5.2.1.3. Infraestructura y recursos recomendados para proyectos de ciencia de datos.....	48
2.5.2.1.4. Herramientas y utilidades recomendadas para la ejecución de proyectos.....	49
2.5.3. KDD (Knowledge Discovery in Databases).....	49

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 7 de 128

2.5.3.1. Integración y recopilación de datos.	50
2.5.3.2. Transformación/Reducción.	50
2.5.3.3. Minería de datos (Datamining).	50
2.5.3.4. Evaluación e interpretación.	51
2.5.3.5. Difusión y uso.	51
2.5.4. OSEM.	51
2.5.4.1. Obtener los datos (O).....	52
2.5.4.2. Limpiar y preparar los datos (S)	52
2.5.4.3. Explorar los datos (E).....	53
2.5.4.4. Modelar los datos (M).....	53
2.5.4.5. Interpretar los resultados (N).....	53
2.5.5. La metodología Canvas	53
2.6. Modelos de detección de objetos.....	56
2.6.1. SSD - Single Shot Detector.	58
2.6.2. Faster R-CNN.	60
2.6.3. Yolo.....	61
2.6.3.1. Arquitectura del modelo Yolo.....	63
2.6.3.2. Evolución del modelo Yolo.	65
2.6.3.2.1. Yolo Versión 1.	65
2.6.3.2.2. Yolo Versión 2.	66
2.6.3.2.3. Yolo Versión 3.	67
2.6.3.2.4. Yolo Versión 4.	69
2.6.3.2.5. Yolo Versión 5.	70
2.6.3.2.6. Yolo Versión 6.	73
2.6.3.2.7. Yolo Versión 7.	74
2.6.3.2.8. Yolo Versión 8.	75
2.7. La transferencia de aprendizaje IA.....	77
2.8. Google colab.....	78
2.8.1. Restricciones.....	80

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 8 de 128

2.9. Dataset	81
2.10. Herramienta Make sense.	82
Capítulo 3. Metodología	83
3.1. Selección de metodología	83
3.1.1. Fases aplicadas al sistema de detección temprana de Varroosis.....	84
3.2. Selección del modelo pre entrenado YoloV5x	86
3.3. Modelo de caso de uso del negocio.....	89
Capítulo 4. Desarrollo e implementación.....	89
4.1. Construcción del conjunto de datos Dataset	90
4.1.1 Características del Conjunto de Entrenamiento	91
4.2. Etiquetado.	92
4.3. Arquitectura del Modelo	94
4.4. Entrenamiento de datos personalizados usando YoloV5x.....	96
4.4.1. Creación de dataset yaml.....	96
4.4.2. Inferencia	98
4.4.3. Entrenamiento.	100
4.4.4. Visualización.....	101
Capítulo 5. Resultados	104
Capítulo 6. Conclusiones y Trabajos Futuros	108
6.1. Conclusiones	108
6.2 Trabajos Futuros.	109
Referencias.....	111

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 9 de 128

Lista de figuras

Imagen 1. Hembra adulta de Varroa.....	27
Imagen 2. Gráfico de un algoritmo de clasificación.....	32
Imagen 3. Gráfico de un algoritmo de regresión.	32
Imagen 4. Diagrama de árbol de decisión.....	34
Imagen 5. Gráfico de redes neurales.	35
Imagen 6. Arquitectura de una Red Neuronal Convolutacional.	38
Imagen 7. Ciclo de vida de minería de datos.....	43
Imagen 8. Data Science Lifecycle.	47
Imagen 9. The Knowledge Discovery in Databases (KDD) Process.	49
Imagen 10. Data Science Project.....	52
Imagen 11. Modelo de detección de objetos en tiempo real.....	57
Imagen 12. Introducción a Yolo: detección de objetos en tiempo real.	62
Imagen 13. Arquitectura del sistema Yolo.....	63
Imagen 14. Operaciones arquitectónicas por capas en el marco Darknet-19.	67
Imagen 15. Arquitectura de Yolo (v3).	68
Imagen 16. Arquitectura de la red YoloV4.	69
Imagen 17. Descripción de los modelos YoloV5.	71
Imagen 18. YoloV5-P5 640.	72
Imagen 19. Arquitectura del modelo YoloV6.....	74
Imagen 20. Proceso de red de marco general de Yolo v7.	75
Imagen 21. Yolov8 Architecture.	76
Imagen 22. Editor Make sense.....	83
Imagen 23. Modelo de caso de uso.....	89
Imagen 24. Foto de Abejas en una colmena.	91
Imagen 25. Etiquetado en MakeSense.	93
Imagen 26. Diagrama de estructura de la red YoloV5x.	95
Imagen 27. Creación de etiquetas para entrenamiento.....	97
Imagen 28. Configuración del cuaderno de entrenamiento.....	97
Imagen 29. Configuración del directorio de datos.....	98
Imagen 30. Referencia comandos de clonación de repositorio YoloV5.	99
Imagen 31. Directorios en Google Colab de YoloV5.....	99
Imagen 32. Comando para descomprimir un archivo.	100
Imagen 33. Visualización del entrenamiento con TensorFlow.....	101
Imagen 34. Exportar modelos.....	101
Imagen 35. Instalación de requerimientos en entorno local PyCharm.....	102
Imagen 36. Personalización del modelo en entorno local PyCharm.....	104
Imagen 37. Entrenamiento de epochs en Google Colab.	105

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 10 de 128

Imagen 38. Métrica de precisión epochs en Google Colab.	106
Imagen 39. Gráfica de Precision-Confidence Curve.....	107
Imagen 40. Resultado detección abeja.....	107

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 11 de 128

Lista de tablas

Tabla 1. Cuadro comparativo Modelos de Machine Learning.	36
Tabla 2. Cuadro comparativo metodologías Machine Learning.	56
Tabla 3. Cuadro comparativo Modelos de detección de imágenes.....	65

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 12 de 128

Glosario

Algoritmo

“Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema”
(Asale, R.-, & Rae. 2023)

Entrenamiento

“Se trata de un procedimiento pensado para obtener conocimientos, habilidades y capacidades.” (Pérez Porto, J., & Merino, M. 2021)

Red pre-entrenada

Una red pre-entrenada es una red neuronal que ha sido previamente entrenada en un conjunto de datos grande y generalmente disponible públicamente. Estas redes se utilizan como punto de partida para aplicaciones específicas y se ajustan o afinan para mejorar su rendimiento en una tarea específica. (Lerch, D. 2021)

Interfaz

“Conexión, física o lógica, entre una computadora y el usuario, un dispositivo periférico o un enlace de comunicaciones” (Asale, R.y Rae. 2023).

Repositorio

Un repositorio es un espacio centralizado donde se almacena, organiza, mantiene y difunde información digital, habitualmente archivos informáticos, que pueden contener trabajos científicos, conjuntos de datos o software. (Wikimedia. 2023)

Modelo de Detección

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 13 de 128

La detección de objetos es un problema de visión informática. Si bien está estrechamente relacionada con la clasificación de imágenes, la detección de objetos realiza la clasificación de imágenes a una escala más granular. La detección de objetos ubica y categoriza entidades dentro de las imágenes. Los modelos de detección de objetos se entrenan normalmente mediante el aprendizaje profundo y las redes neuronales. (Luis Quintanilla. 2023)

Apis mellifera

Conocida como abeja de la miel europea, es una especie de insecto perteneciente al orden *Hymenoptera* y a la familia *Apidae*. Es la especie de abeja con mayor distribución global, siendo originaria de Europa y África, el noroeste de Asia, y expandiéndose a América y Australia por acciones antrópicas. (“*Apis mellifera*: características,”. 2020)

Varroa destructor

“La *Varroa destructor* de abejas, ácaro parasitario, es una de las plagas más graves de la abeja melífera, *Apis mellifera*”. (*Varroa destructor*: USDA ARS. 2023)

Polinización

“Proceso mediante el cual el grano de polen llega al estigma de una flor.” (Asale, R. y Rae. 2023).

La polinización es el transporte del polen de las plantas desde los estambres (órganos masculinos de las flores que originan los granos de polen) hasta la parte receptiva de las flores, lugar donde germina y fecunda los óvulos de la flor, haciendo posible la producción de semillas y frutos. (Significados. Polinización. 2016)

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 14 de 128

Capítulo 1. Introducción.

En la apicultura mundial *Varroa destructor*, ha causado un impacto negativo durante varios años en los cuales ha tomado mayor fuerza y se ha propagado hacia la mayoría de los países productores de miel (Grosso, Patricia y Arias, 2013). Algunos de los países que se han visto afectados han sido los siguientes Estados Unidos, países de Europa, Asia, África y América Latina, pero especialmente atacando con mayor severidad a China y Rusia Según (Rosenkranz et al., 2010) logrando así convertirse en una gran amenaza para la salud de las abejas y también para los apicultores.

La lucha contra esta enfermedad se ha convertido en un gran desafío para la apicultura, ya que la disminución de este proceso podría generar grandes consecuencias en cuestión de producción alimentaria afectando hasta los lugares más lejanos (María & Ferrari, 2016). En países como México, hoy en día se han establecido medidas de control obligatorias en las cuales el objetivo principal es tratar de utilizar los métodos menos dañinos para controlar la plaga prevaleciendo la conservación del medio. (SENASICA. 2021).

A nivel Nacional la apicultura es una práctica importante para el bienestar de las comunidades locales debido a que es una actividad productiva que ofrece múltiples beneficios tanto a los apicultores como a los ecosistemas y demás atributos de la biodiversidad. (Olaya, M. 2017). La afectación de esta enfermedad ha tomado diferentes comportamientos en base a los múltiples factores a los que se encuentran expuestos principalmente el clima ya que Colombia es un país con diferentes altitudes, latitudes y topografías lo cual genera que los climas varían

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 15 de 128

dependiendo de la región en la que se encuentre, convirtiéndose así en un gran foco de atracción de enfermedades como la Varroosis. (Grosso, Patricia, y Arias, 2013)

Considerando que en Colombia la Varroosis es un problema de reciente introducción, no existe conocimiento en aspectos básicos de la biología y etología en el medio, sumado a lo anterior los apicultores desconocen por completo las medidas de prevención y control, aplicando muchas veces métodos irracionales para tal fin. (Efrén y Jorge, 2016).

A nivel local el Departamento del Cauca es una de las regiones con gran producción en productos relacionados a la apicultura, sin embargo, el clima ha resultado ser el enemigo número uno debido a que este es principalmente cálido y húmedo propicio para la propagación de diferentes tipos de enfermedades entre ellas la Varroosis, la cual aprovecha este tipo de condiciones para lograr propagarse rápidamente. (“En el Cauca, el ICA,” 2022).

Adicional a esto, es importante conocer que el clima no es la única afectación en esta área, también se encuentra afectada por el manejo poco adecuado de las diferentes colmenas debido a que los apicultores deben estar vigilando el mayor tiempo posible las colmenas, trabajo que a veces demanda mucho tiempo y que no es posible realizar. (Bounous, C., y Boga, V. 2005)

Es por esto que los índices de contagio cada vez más se encuentran en mayor incremento y la apicultura toma mayor importancia (Klatt et al., 2019). Sin embargo, su supervivencia está amenazada por diversas causas, como el cambio climático y la agricultura intensiva (Potts et al., 2010).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 16 de 128

1.1. Determinar la *Varroa destructor* en colonias de abejas del Macizo Colombiano.

Se realizó una encuesta (ver Anexo A), para determinar la presencia de *Varroa destructor* en colonias de abejas del Macizo Colombiano, ya que esta especie de ácaro puede ser una de las principales causas de pérdida de colonias de abejas en la región. La Varroosis es una enfermedad que debilita a las abejas y puede transmitir otros patógenos, afectando no solo a las abejas, sino también a la polinización y la producción de miel. (Grosso, G. S. 2014)

La encuesta se aplicó a 144 apicultores y conocedores del tema con el objetivo de determinar la existencia de la *Varroa destructor* en colonias de abejas del Macizo Colombiano. Las preguntas más relevantes de la encuesta se centran en la presencia de Varroosis y los efectos que puede tener sobre las abejas y su producción.

Por medio del diálogo y la observación directa con los apicultores se aprendió a reconocer la Varroa. Se observaron las celdas en las colmenas y se identificó la Varroa sobre la abeja. Los apicultores señalaron cuáles son los colores que las identifican y el tamaño que tienen.

Estas preguntas son importantes en la encuesta porque proporcionan información relevante para la investigación. Aquí se explica por qué son importantes:

A. La Varroosis es una enfermedad parasitaria que afecta a las abejas, causada por el ácaro *Varroa destructor*. ¿En su apiario ha tenido presencia de esta enfermedad?

- Si 84%
- No 16%

Esta pregunta es importante porque proporciona una visión general de la incidencia de la enfermedad en el área específica del Macizo Colombiano y permite evaluar el alcance del

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 17 de 128

problema. Esta información es esencial para comprender la magnitud de la infestación de *Varroa destructor* y su impacto en las colonias de abejas en la región. Además, al conocer la presencia de la enfermedad en los apiarios, se pueden identificar áreas específicas que requieren una atención especial en términos de prevención y control de la Varroasis.

Los resultados obtenidos fueron que el 84% de la población encuestada respondieron que sí han tenido presencia de esta enfermedad en sus apiarios (ver Anexo A).

B: ¿Con qué frecuencia realizas tratamientos contra la Varroasis en tus colonias de abejas?

- Una vez al año 23.6%
- Dos veces al año 15.3%
- Tres veces al año 38.9%
- Más de tres veces al año 19.4%
- No realizo tratamientos 10.4%

Esta es una de las preguntas más importantes porque permite identificar si los apicultores están siguiendo protocolos de tratamiento regulares y adecuados. Además, proporciona información sobre la efectividad de los métodos utilizados. Si la frecuencia de tratamiento es baja o inexistente, esto podría indicar un mayor riesgo de presencia y propagación de *Varroa destructor* en las colonias de abejas del Macizo Colombiano.

- Los resultados obtenidos son que 38.9% de las personas encuestadas realizan tratamientos 3 veces al año (ver Anexo A).

C. La pregunta "¿Has notado una disminución en la producción de miel o en la salud de tus colonias de abejas debido a la Varroasis?"

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 18 de 128

- sí 70,1%
- no 10.4%
- no lo sé 20.8%

Esta pregunta es muy importante porque permite obtener información relevante sobre los impactos de la Varroasis en la producción de miel y la salud de las colonias de abejas (ver Anexo A).

Basándonos en los resultados de la encuesta, donde el 70.1% de los apicultores afirma haber notado una disminución en la producción de miel o en la salud de sus colonias de abejas debido a la Varroasis, podemos concluir lo siguiente:

Que el Impacto es significativo: La Varroasis está teniendo un impacto importante en la producción de miel y la salud de las colonias de abejas, ya que una mayoría significativa de los apicultores ha experimentado una disminución en estos aspectos.

Durante la aplicación de la encuesta se conoció la varroa, y los apicultores nos enseñaron como la detectan pues, es un punto generalmente de color rojo naranja o negruzco según el ciclo de vida de la varroa y la abeja, se definió para nuestro estudio que se identificarían sobre el cuerpo de las abejas adultas dicho punto.

1.2. Propuesta

Se busca desarrollar un sistema informático que analiza imágenes y al mismo tiempo identifique la Varroosis en las abejas a través de un sensor, que a futuro puede tener uso, al instalarlos en la entrada de los apiarios. Para cumplir este propósito es necesario y de la Varroosis,

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 19 de 128

de manera que las imágenes que se capturen por las cámaras serán comparadas con el conjunto de datos y se podrá detectar la Varroa. Este sistema busca implementar un algoritmo que utilice el conjunto de datos, de manera que más adelante permita que el apicultor pueda tomar decisiones a tiempo y así evitar la propagación de la enfermedad.

Este trabajo traerá diferentes beneficios en el área de la apicultura. Primero, al crear el sistema de detección temprana de la Varroosis permitirá tomar medidas a tiempo para evitar la muerte de las abejas. Como segundo beneficio la detección a tiempo la Varroa, reducen los costos de producción, permite un ahorro de tiempo y recursos para los apicultores y a su vez implica la implementación de nuevas tecnologías que mejoran la eficiencia en todo el proceso.

Disponer un conjunto de datos con imágenes obtenidas en diferentes apiarios y en varias etapas de desarrollo de las abejas, facilita la identificación de la Varroosis y mejorar la productividad de la colmena

Adicional a ello se puede tomar el Dataset como referencia para próximas propuestas e investigaciones en el campo de la apicultura. Otro beneficio, es la conservación del medio ambiente, ya que la detección de Varroa a tiempo disminuye la mortalidad de las abejas, lo que permite la polinización de los diferentes cultivos.

Capítulo 2. Marco Teórico

En este marco teórico se explorará inicialmente las abejas y enfermedades como la Varroosis la cual tiene graves afectaciones, después se hizo énfasis principalmente en cómo la inteligencia artificial, específicamente el *Machine Learning*, puede ayudar a detectar

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 20 de 128

tempranamente la Varroosis en las abejas. y los diferentes tipos de aprendizaje, modelos y fases del proceso de aprendizaje, así como las metodologías más utilizadas en este campo y algunos modelos de detección de objetos. Se analizarán los desafíos asociados con la detección temprana de la Varroosis y cómo la inteligencia artificial puede contribuir a abordarlos.

El objetivo final es proporcionar una visión general de cómo la inteligencia artificial puede mejorar la salud de las colmenas y prevenir la propagación de esta enfermedad en las abejas.

2.1. Las Abejas

Según Michener (1974), “Las abejas son un grupo de insectos visitantes de flores que abandonaron sus hábitos de avispa de aprovisionar sus nidos con insectos o arañas y en cambio alimentan a sus larvas con polen y néctar recolectado de flores o con secreciones glandulares, finalmente derivadas de la mismas fuente”.

Las abejas son unos insectos extremadamente sociables que viven en colonias que se establecen en forma de enjambres y en los que se organizan en una estricta jerarquía de tres rangos sociales: la abeja reina, los zánganos y las abejas obreras.

Las obreras son hembras que realizan diversas tareas en la colmena, como buscar comida, construir y proteger la colmena. La reina tiene la función de poner huevos y regular las actividades de la colmena, mientras que los zánganos son los machos que viven en la colmena durante la primavera y el verano. Cada una de estas abejas desempeñan un papel crucial en la supervivencia y funcionamiento de la colmena. (Nates-Parra, 2011)

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 21 de 128

Estos insectos habitan en todos los continentes de la Tierra excepto en la Antártida, y se trata de uno de los insectos más antiguos, del que se sabe, puebla nuestro planeta desde hace más de 30 millones de años. (Nates-Parra, 2011)

2.1.1. Características

En general, las abejas son más robustas y peludas que las avispas, aunque algunas de ellas presentan un fenotipo vespoideo (cuerpo delgado, largo y con poca pilosidad; p. ej. *Hylaeus*, nómadas y en general abejas parasíticas) (“Abejas - Características,” 2023)

Algunas características morfológicas ayudan a diferenciarlas (Michener et al. 1994):

1. Presencia de pelos plumosos o ramificados en varias partes del cuerpo y de las patas, aunque algunas veces están restringidos a pocas áreas, p. ej. El propodeo. (Equipo de redacción, 2022)
2. El basitarso posterior es más ancho que los segmentos siguientes del tarso y generalmente es aplanado (Equipo de redacción, 2022).

Los *esfeciformes* (una de las divisiones de *Apoidea* que agrupa a las avispas *esfecoideas*) tienen el primer y segundo segmentos tarsales similares en ancho y junto con el espolón tibial forman un órgano limpiador de antenas (*strigil*) (Equipo de redacción, 2022).

Ausencia de pelos dorados o plateados en la parte inferior de la cara. Esta característica es especialmente importante cuando se trata de separar abejas con fenotipo *vespoideo* de muchas avispas *esfecoideas*: éstas presentan pelos dorados o plateados que, a la luz, les confieren brillo en la cara (Equipo de redacción, 2022).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 22 de 128

2.1.1.1. *Apis mellifera*

Se puede encontrar desde especies parásitas en nidos de otras abejas hasta formas cuasisociales; esto es, pequeñas colonias en las que dos o más hembras construyen, aprovisionan y ovipositan las celdas cooperativamente; generalmente son de la misma edad y la misma generación o comunales. Dentro de esta familia se encuentra la *Apis mellifera* la especie de abeja más utilizada por el ser humano en polinización, aunque no la única, es *Apis mellifera*, principalmente por sus hábitos generalistas de forrajeo y su facilidad de manejo. Por su producción de miel, esta especie, de origen africano y sudeuropeo, fue domesticada e introducida en casi todo el mundo (“Honeybee.jpg,” 2022).

Nombre común: Abeja Melífera (también conocida como abeja europea o doméstica) (“Honeybee.jpg,” 2022)

Nombre científico: *Apis mellifera* (“Honeybee.jpg,” 2022).

Tipo: Invertebrados (“Honeybee.jpg,” 2022).

Dieta: Herbívoro (“Honeybee.jpg,” 2022).

Nombre del grupo: Colonia, enjambre (“Honeybee.jpg,” 2022).

Promedio de vida en la naturaleza: Hasta 5 años (“Honeybee.jpg,” 2022).

Tamaño: 1 a 1,5 centímetros (obreras) (“Honeybee.jpg,” 2022).

2.1.2. Importancia.

Las abejas son esenciales para la reproducción de plantas, la producción de alimentos, la biodiversidad, la producción de productos naturales y el equilibrio de los ecosistemas. Su

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 23 de 128

importancia radica en su función de polinizadores y en los servicios eco sistémicos que brindan, lo que impacta directamente en nuestra calidad de vida y en la salud del planeta.

2.2. La Varroosis en las abejas

La Varroa ha sido reportada como un coadyuvante de la disminución de las abejas. Se comporta como parásito de abejas adultas y de sus crías, al ser parasitadas desde ese estado larvario, las abejas nacen ya con serios problemas de salud como, por ejemplo, las abejas adultas muestran varios cambios de comportamiento y se acorta considerablemente su vida. (Cánovas Alcázar, 2006)

2.2.1. Etiología.

Varroa jacobsoni o *destructor*: fácil identificación, parecido a las garrapatas.

Color rojizo, mide entre 1,7 a 1,8 mm de ancho por 1,1 a 1,2 mm de largo.

Succiona hemolinfa de larvas y de abejas adultas. (Jong, 1983; Ritter, 1996).

2.2.2. Historia

Originariamente *Varroa jacobsoni* es un parásito de la abeja *Apis* utilizada en Asia para la producción de miel donde el equilibrio entre parásito y hospedador no hace necesaria la realización de tratamientos. Sin embargo, cuando *Apis mellifera* (nuestra abeja de la miel) fue introducida en Asia para obtener mejores rendimientos fue cuando Varroa se convirtió en un parásito produciendo graves daños, hasta el punto de acabar con las colonias que no eran tratadas (Cánovas Alcázar, 2006). A lo largo del siglo XX, Varroa se ha ido extendiendo prácticamente por todo el mundo. En España fue detectada a mediados de la década de los ochenta, encontrando actualmente en todo el territorio nacional, excepto en la isla de La Palma. En el año 1993 se

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 24 de 128

identificó por primera vez en Colombia, la presencia de la *Varroa jacobsoni* en el Municipio de Fusagasugá (Cundinamarca). La Varroosis la produce el ácaro ectoparásito *Varroa destructor* que afecta a las crías y a las abejas adultas, de las que se alimenta expoliando su hemolinfa (Cánovas Alcázar, 2006).

De acuerdo con el VII Congreso SEAE Zaragoza 2006 N° 6, su aparición ha provocado una mortandad masiva de colmenas y la desaparición de gran número de apicultores (Cánovas Alcázar, 2006).

2.2.3. Morfología

Morfológicamente la Varroa presenta la estructura típica de los ácaros presentando diferencias entre la hembra y el macho (Chávez y García, 1993). La hembra tiene el cuerpo ovalado de color castaño-rojizo. Mide 1,1 mm de largo por 1,6 mm de ancho y se ve a simple vista. Vive sobre el cuerpo de la abeja y se alimenta de su sangre (hemolinfa). Tiene el dorso abombado y patas relativamente cortas y encorvadas, dispuestas de tal forma que el primer par se asemeja a unas antenas (Chávez y García, 1993).

Por el contrario, el macho es más pequeño y de color blanco, vive dentro de las celdillas operculadas (cubiertas por una cúpula de cera, bajo la cual se producirá la metamorfosis de la abeja), muriendo después de la fecundación. No extrae sangre a la abeja y su misión en el ciclo es la de aparearse con las hembras (Chávez y García, 1993).

2.2.4. Ciclo

En el ciclo de vida de la Varroa se debe distinguir una fase reproductiva, dentro de las celdillas de cría operculadas y una fase forética sobre las abejas adultas. En ambas fases el parásito

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 25 de 128

se va a alimentar succionando la hemolinfa del hospedador, consecuencia de lo cual va a generar graves daños en las abejas (Cánovas Alcázar, 2006). En la fase reproductiva las hembras de Varroa eligen una celda que contenga una larva finalizando su etapa abierta, teniendo preferencia por las de zánganos. Esta preferencia es debida a un mayor tamaño de las celdillas (que proporciona más alimento y por tanto deposita un mayor número de huevos) y una menor temperatura de estas celdas (al encontrarse en los bordes del panal, lo que facilita la multiplicación del parásito) (Cánovas Alcázar, 2006).

Las hembras de Varroa fecundadas ponen entre tres y cinco huevos si están en una celda de obrera y entre tres y siete si están en una de zángano. El periodo de desarrollo del huevo es de cinco a siete días, en el caso del macho y de siete a nueve días, en el de la hembra. A partir de una celdilla de obrera se obtiene una descendencia de 1,5 a 2 hembras de Varroa y de 2,5 a 3 cuando se parte de una celdilla de zángano (Cánovas Alcázar, 2006).

Una vez que la joven abeja abandona la celdilla de cría, no puede deshacerse de la Varroa, ya que ésta escapa a los movimientos de cepillado (Cánovas Alcázar, 2006).

2.2.5. Daños

La infestación de las colmenas de una explotación por el ácaro *Varroa jacobsoni* suelen presentar variaciones atribuibles a diversas condiciones como el estrés, la edad de las reinas, las reservas de alimento y las características genéticas. (Cánovas Alcázar, 2006). Su forma de vida causa dos tipos de daños:

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 26 de 128

2.2.5.1. Daños directos sobre la cría de las abejas.

Causan malformaciones durante el desarrollo, generando abejas de menor tamaño, cuerpo deforme y alas atróficas, lo que las invalida para el desarrollo de sus funciones en la colonia y acaban por ser eliminadas. Los daños causados en las abejas adultas durante la fase forética son menos aparentes, pero igualmente influyen en el desarrollo de las actividades de las abejas (Flores et al., 2000).

2.2.5.2. Daños indirectos.

La transmisión de enfermedades, especialmente cuadros víricos, que probablemente sean los responsables de las pérdidas de parte de la colonia achacadas al parásito (Cánovas Alcázar, 2006). De acuerdo con el VII Congreso SEAE Zaragoza 2006 N° 6, evidentemente, las mermas producidas en la población de abejas se traducirán en disminución de la producción y, con demasiada frecuencia, en la muerte de las colmenas. (Cánovas Alcázar, 2006).

2.2.6. Modo de Propagación

Las abejas obreras y los zánganos suelen ser el vehículo para la propagación del parásito entre colmenas (Cánovas Alcázar, 2006). Generalmente las abejas obreras reconocen su colmena al volver de sus vuelos de pecoreo (recolección). Sin embargo, en algunas ocasiones entran en otras colmenas, donde son aceptadas por llevar la carga recolectada. En otros casos, estas abejas acuden a otras colmenas débiles para robar la miel (Cánovas Alcázar, 2006). El caso de los zánganos es más simple, por tener libre acceso a cualquier colmena durante la temporada reproductiva (primavera-verano) (Cánovas Alcázar, 2006). Otras formas de propagación de la

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 27 de 128

Varroa son la manipulación descuidada del apicultor, la trashumancia no controlada de colmenas, así como la no detección precoz de la enfermedad (Cánovas Alcázar, 2006).

2.2.7. *Varroa destructor*

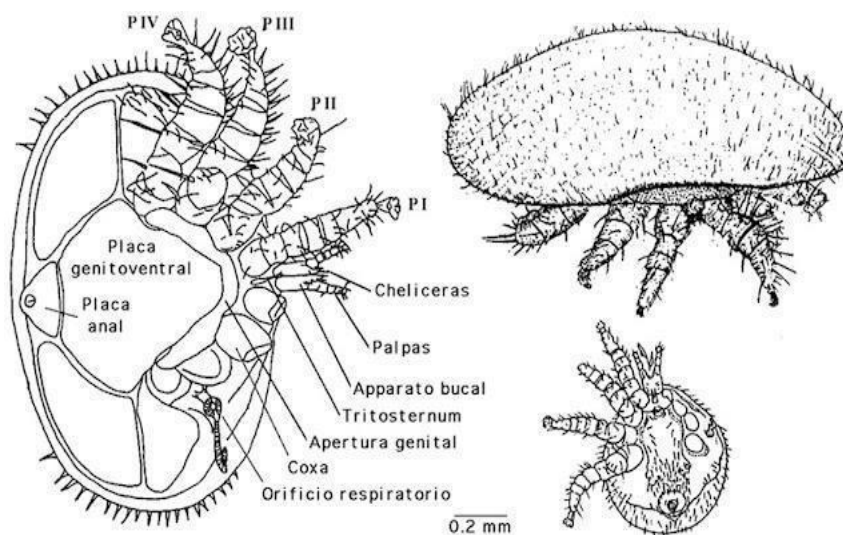


Imagen 1. Hembra adulta de Varroa.

La Familia de la Apicultura - (The Beekeeping Family, 2023).

2.2.8. Métodos de control más reconocidos

2.2.8.1. *Tratamientos con productos naturales.*

Según el Reglamento (CEE) N° 2092/91 de fecha 24 de junio de 1991 sobre la producción agrícola ecológica y su indicación en los productos agrarios y alimenticios, “en los casos de infestación por *Varroa jacobsoni* podrán utilizarse los siguientes ácidos: fórmico, láctico, acético y oxálico, así como las siguientes sustancias: mentol, thimol, eucalyptol o alcanfor”.

En aquellas sustancias que actúan por evaporación, la liberación del principio activo se encuentra mediatizado por la temperatura interior de la colmena (que a su vez fluctúa dependiendo

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 28 de 128

de la temperatura exterior) y por la naturaleza y dimensiones de la superficie de evaporación del dosificador utilizado (Cánovas Alcázar, 2006). El principal problema que plantea su uso es la falta de control en la evaporación de los mismos, lo que ha causado en ocasiones graves problemas en los colmenares tratados (abandono de las colmenas por las abejas) (Cánovas Alcázar, 2006).

Por otro lado, en zonas apícolas es frecuente que a partir de colmenares infestados no tratados se produzca la re-infestación de los que sí lo han sido, de tal manera que en pocos meses una colmena pueda re infestarse y aumentar a tal velocidad la población de parásitos que se convierta en una amenaza para su supervivencia, lo que obliga a la aplicación de tratamientos con mayor frecuencia (Cánovas Alcázar, 2006). Con el fin de eliminar la mayor parte de los ácaros de las colmenas y que éstas puedan sobrevivir durante el invierno, en determinadas zonas apícolas, donde la presencia de cría de abejas se extiende durante todo el año, es necesario realizar dos tratamientos, uno a comienzos de la estación (cuando las abejas salen de su período de reposo) y otro a fin de temporada (post-cosecha).

2.2.8.2. Técnicas culturales Colmenas control.

Mediante la utilización de fondos sanitarios en apicultura ecológica, donde las condiciones meteorológicas o el grado de infestación de la colmena hacen que la eficacia de los tratamientos realizados no sea tan efectiva como desearíamos, es conveniente la utilización de fondos sanitarios. VII Congreso SEAE Zaragoza 2006 Nº 6 Éstos, nos ayudarán a evaluar el grado de infestación de las colmenas y a programar los tratamientos que debamos realizar (Cánovas Alcázar, 2006).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 29 de 128

Los fondos sanitarios consisten en adaptar a la colmena un doble fondo en el que pondremos una lámina de plástico rígido blanco impregnado de vaselina filante. Los ácaros de la Varroa, cuando mueren de forma natural o por la acción de alguno de los tratamientos realizados, caen a esta lámina de plástico quedando pegados en la vaselina (Cánovas Alcázar, 2006). Periódicamente, retiraremos la lámina y procederemos al conteo de los ácaros. Hay que tener en cuenta que los ácaros de Varroa son de color oscuro. La lámina debe de estar cuadriculada, con el fin de poder realizar este conteo de forma más cómoda. La periodicidad de conteo de ácaros no es conveniente que sea superior a 2 ó 3 días, puesto que con más tiempo la excesiva acumulación de residuos dificultaría el conteo (Cánovas Alcázar, 2006).

Una vez realizado el conteo, podremos calcular la mortalidad diaria dividiendo el número de ácaros por los días que la lámina ha estado en el fondo de la colmena. De esta forma podremos realizar un estudio sobre la evolución de la plaga en nuestro colmenar y determinar los crecimientos de población del ácaro y el momento óptimo para la realización de los tratamientos (Cánovas Alcázar, 2006).

Aunque lo ideal sería tener estos fondos sanitarios en todas las colmenas, para determinar la evolución de la plaga en el colmenar y por comodidad para el apicultor, se pueden poner en, aproximadamente, el 10% de las colmenas. Es conveniente que estas colmenas control se encuentren distribuidas o repartidas por todo el colmenar, con el fin de poder establecer una media del grado de infestación del colmenar.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 30 de 128

2.3. Inteligencia artificial

La inteligencia artificial es un campo de la informática que se enfoca en la creación de sistemas y programas que pueden realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, la toma de decisiones y el reconocimiento de patrones. (2023, March 23). Se basa en algoritmos y modelos matemáticos que permiten a las máquinas aprender de los datos y mejorar su rendimiento con el tiempo. (Landeck et al., 2020, p. 165).

Se divide en dos categorías principales: la inteligencia artificial débil y la inteligencia artificial fuerte. La inteligencia artificial débil se refiere a sistemas que están diseñados para realizar tareas específicas, como el reconocimiento de voz o la detección de fraudes. La inteligencia artificial fuerte, por otro lado, se refiere a sistemas que pueden realizar cualquier tarea que un ser humano pueda hacer. ("Inteligencia Artificial," 2017)." La inteligencia artificial es una tecnología que se utiliza cada vez más en la apicultura para mejorar la gestión de las colmenas" (Landeck et al., 2020, p. 165).

2.4. El Machine Learning

Es una rama de la Inteligencia Artificial que estudia sistemas capaces de aprender a realizar una tarea a partir de datos de ejemplo. Es de naturaleza inductiva, a diferencia de la inteligencia artificial clásica, y comprende técnicas y métodos para realizar clasificación, optimización y predicción, mayormente en dominios en donde los problemas no pueden definirse de forma explícita o no existen soluciones analíticas aplicables. (Sandoval Serrano, L. J. 2018).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 31 de 128

2.4.1. Tipos de aprendizaje.

2.4.1.1. Aprendizaje supervisado.

Es cuando se entrena un algoritmo de Machine Learning dándole las preguntas (características) y las respuestas (etiquetas). Así en un futuro el algoritmo pueda hacer una predicción conociendo las características. En este tipo de aprendizaje hay dos algoritmos (entrenamientos): el de clasificación y el de regresión. (Sandoval Serrano, L. J. 2018).

- **Algoritmo de clasificación.**

Se espera que el algoritmo indique a qué grupo pertenece el elemento en estudio. El algoritmo encuentra patrones en los datos que se le dan y los clasifica en grupos. Luego compara los nuevos datos y los ubica en uno de los grupos y es así como puede predecir de qué se trata. La variable por predecir es un conjunto de estados discretos o categóricos. Pueden ser: Binaria: {Sí, No}, {Azul, Rojo}, {Fuga, No Fuga}, etc. Múltiple: Comprará {Producto 1, Producto 2...}, etc. Ordenada: Riesgo {Bajo, Medio, Alto}, etc. (Sandoval Serrano, L. J. 2018).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 32 de 128

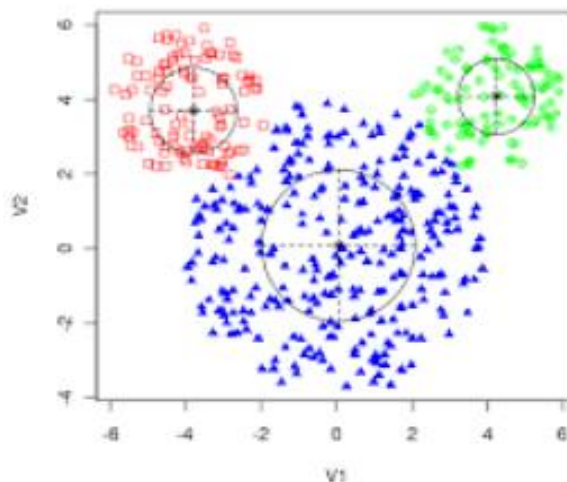


Imagen 2. Gráfico de un algoritmo de clasificación.

Sandoval Serrano, L. J. (2018).

- **Algoritmo de regresión.**

En este método lo que se espera es un número. No lo ubica en un grupo, sino que devuelve un valor específico. (Sandoval Serrano, L. J. 2018).

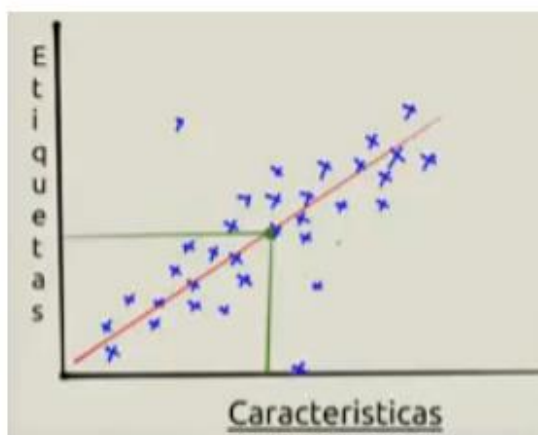


Imagen 3. Gráfico de un algoritmo de regresión.

Sandoval Serrano, L. J. (2018).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 33 de 128

2.4.1.2. Aprendizaje No supervisado.

Aquí se asignan las características al algoritmo, nunca las etiquetas. lo que se busca es que agrupe los datos dados según sus características. El algoritmo solo sabe que como los datos comparten ciertas características, de esa forma asume que pueda que pertenezcan al mismo grupo. (Sandoval Serrano, L. J. 2018).

2.4.2. Modelos de Machine Learning.

Los algoritmos de *Machine learning*, se pueden agrupar en tres modelos:

2.4.2.1. Modelos lineales.

Estos tratan de encontrar una línea que se “ajuste” bien a la nube de puntos que se disponen. Aquí destacan desde modelos muy conocidos y usados como la regresión lineal (también conocida como la regresión de mínimos cuadrados), la logística (adaptación de la lineal a problemas de clasificación -cuando son variables discretas o categóricas) (Sandoval Serrano, L. J. 2018).

Estos dos modelos tienen el problema del “*overfit*”, esto significa que se ajustan “demasiado” a los datos disponibles, con el riesgo que esto tiene para nuevos datos que pudieran llegar. Al ser modelos relativamente simples, no ofrecen resultados muy buenos para comportamientos más complicados (Sandoval Serrano, L. J. 2018).

2.4.2.2. Modelos de árbol.

Son modelos precisos, estables y más sencillos de interpretar básicamente porque construyen unas reglas de decisión que se pueden representar como un árbol (Sandoval Serrano,

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 34 de 128

L. J. 2018). A diferencia de los modelos lineales, pueden representar relaciones no lineales para resolver problemas. En estos modelos, destacan los árboles de decisión y los *random forest* (una media de árboles de decisión). Al ser más precisos y elaborados, obviamente ganamos en capacidad predictiva, pero perdemos en rendimiento (Sandoval Serrano, L. J. 2018).

Un árbol de decisión es un algoritmo de aprendizaje supervisado no paramétrico, que se utiliza tanto para tareas de clasificación como de regresión. Tiene una estructura de árbol jerárquica, que consta de un nodo raíz, ramas, nodos internos y nodos hoja (“¿Qué es un árbol,” IBM? 2023).

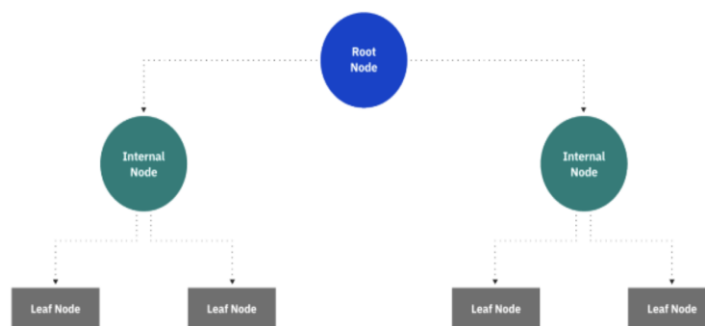


Imagen 4. Diagrama de árbol de decisión.

IBM. (2023, May 19)

2.4.2.3. Redes neuronales.

Las redes artificiales de neuronas tratan, en cierto modo, de replicar el comportamiento del cerebro, donde tenemos millones de neuronas que se interconectan en red para enviar mensajes unas a otras. (Sandoval Serrano, L. J. 2018). Esta réplica del funcionamiento del cerebro humano

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 35 de 128

es uno de los “modelos de moda” por las habilidades cognitivas de razonamiento que adquieren. El reconocimiento de imágenes o vídeos, por ejemplo, es un mecanismo complejo y una red neuronal es lo mejor para realizarlo. (Sandoval Serrano, L. J. 2018).

El problema, como ocurre con el cerebro humano, es que son lentas de entrenar y necesitan mucha capacidad de cómputo. Quizás sea uno de los modelos que más ha ganado con la “revolución de los datos”. (Sandoval Serrano, L. J. 2018).

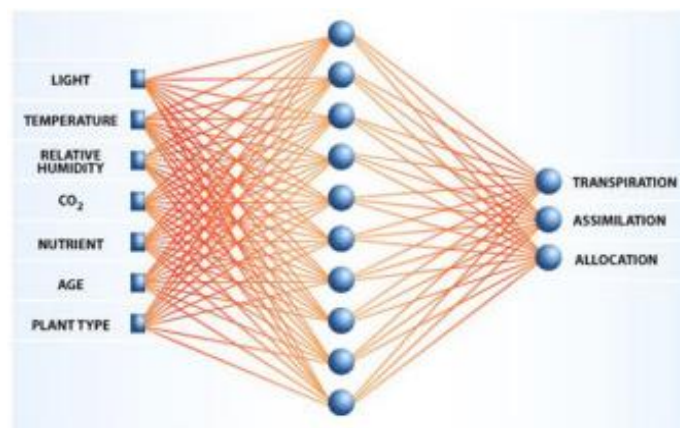


Imagen 5. Gráfico de redes neurales.

Sandoval Serrano, L. J. (2018).

2.4.3. Cuadro comparativo de los modelos de Machine Learning.

De acuerdo con la información de los diferentes modelos de *Machine learning* mencionados anteriormente, se realiza el siguiente análisis que permite contrastar las características y datos de cada una de ellas.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 36 de 128

MODEL O	VENTAJAS	DESVENTAJAS	APLICACIONES
Regresi ón Lineal	Pueden utilizarse para simular una amplia gama de relaciones, y debido a su forma, que es más simple que la de las redes neuronales, sus parámetros estadísticos se analizan y comparan con facilidad, lo que permite que se les extraiga información valiosa. (Saavedra, J. A. 2023) La regresión lineal sigue siendo una buena opción cuando se desee un modelo simple para una tarea de predicción básica. La regresión lineal también suele funcionar bien en conjuntos de datos muy dimensionales, dispersos y sin complejidad. (likebupt. 2023, May 21)	No se puede modelar relaciones complejas. No se pueden capturar relaciones no lineales sin transformar la entrada, por lo que tienes que trabajar duro para que se ajuste a funciones no lineales. (Gonzalez, L. 2020) Los valores atípicos pueden tener efectos enormes en la regresión y los límites son lineales en esta técnica. (ML: ventajas y desventajas – Barcelona Geeks. 2023)	Se utiliza en los campos científicos y en los negocios, y en las últimas décadas se ha utilizado en el aprendizaje automático. (Saavedra, J. A. 2023) De acuerdo con Luis Pardo (2022) en su post de Platzi, tiene aplicaciones en la economía, finanzas, modelos de series de tiempo, ciencia de datos, predicciones sobre una variable numérica, entre otros. Por otra parte, Xabier Carrera (2023) en el tutorial de regresión lineal, indica que se tiene aplicaciones en Estadística, álgebra lineal y alta dimensionalidad.
Árboles de decisió n	Posee una precisión (accuracy en inglés) regular, mejor que los métodos tradicionales o los mencionados anteriormente, y logra tener una explicabilidad ante organismos reguladores (Dumitrescu, Hué, Hurlin, & Tokpavi, 2021). Obtuvo mayor precisión y funcionó más rápido frente a SVM y KNN (Teng & Lee, 2020). Fácil de interpretar: La lógica booleana y las representaciones visuales de los árboles de decisión los hacen más fáciles de entender y consumir. La naturaleza jerárquica de un árbol de decisión también facilita ver qué atributos son los más importantes, lo que no siempre es claro con otros algoritmos, como las Redes neuronales. (“¿Qué es un árbol,” IBM. 2023)	Los árboles complejos son difíciles de interpretar. Es posible la duplicación dentro del mismo subárbol. (González, L. 2020) Más costoso: Dado que los árboles de decisión adoptan un enfoque de búsqueda codicioso durante la construcción, pueden ser más costosos de entrenar en comparación con otros algoritmos. Propenso al sobreajuste: Los árboles de decisión complejos tienden a sobre ajustarse y no se generalizan bien a los nuevos datos. (“¿Qué es un árbol,” IBM. 2023) Inestable: es importante asegurarse de que los valores en el árbol de decisiones sean precisos para que tus ecuaciones sean más exactas. Si cambias incluso una pequeña parte de los datos, los datos más grandes pueden presentar errores. (Asana. 2023)	Los árboles de decisión se emplean generalmente en Big Data para predecir la probabilidad de conseguir un determinado resultado en base a ciertas condiciones (incertidumbre). Ejemplos típicos de la utilización de este tipo de algoritmos son: Hacer estimaciones de las primas de seguros a cobrar a los asegurados. Predecir si se le debe ofrecer un determinado producto a una persona. (“Árboles de decisión,” 2023)
Redes Neuron ales	Tiene la capacidad propia de encontrar patrones y no se necesita de etiquetas previas (Mashrur, Luo, Zaidi, & Robles-Kelly, 2020). Es adecuada en los problemas para los que un modelo de regresión más tradicional no puede encontrar una solución. (likebupt. 2023)	El modelo puede sufrir dificultades de generalización y con ello producir modelos donde exista sobreajuste. (Zhang, 2020, p. 617).	Clasificación de imágenes y visión computacional Aplicaciones de lenguaje natural y reconocimiento de voz. (“¿Qué es Deep,” IBM. 2023) Aplicaciones a la ingeniería biomédica: - Ómica, - Imagenología

Tabla 1. Cuadro comparativo Modelos de Machine Learning.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 37 de 128

2.4.4. Redes Neuronales Convolucionales (CNN)

Son aquellas que constituyen actualmente el estado del arte de varios problemas de visión computacional, dado su buen desempeño en problemas de reconocimiento e interpretación en imágenes y video. Su capacidad para actuar adecuadamente en estos contextos está basada en características fundamentales: conexiones locales, pesos compartidos, *Pooling* y el uso de una gran cantidad de capas (Massiris, Delrieux, y Fernández, 2020).

El propósito de CNN es extraer todas las características de una imagen y luego usar dichas características para detectar o clasificar los objetos en una imagen. Los parámetros de los filtros que se pueden aprender en estas capas; se ajustarán y optimizarán junto con los componentes de clasificación para minimizar el error de clasificación total (Massiris, Delrieux, y Fernández, 2020).

La aparición de CNN ha llevado a un rápido desarrollo del campo de detección de objetos. Por esto nos permitimos mencionar los algoritmos avanzados y más recientes en el campo de detección de objetos con redes neuronales convolucionales: *Faster R-CNN*, *Single Shot Multibox Detector* (SSD) y *You Only Look Once* (Yolo), que es la adoptada en este trabajo y se describe en detalle a continuación (Massiris, Delrieux, y Fernández, 2020).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 38 de 128

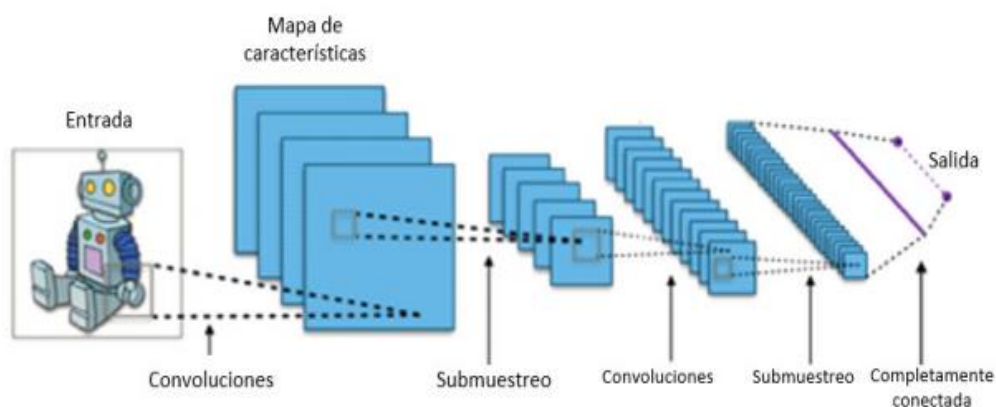


Imagen 6. Arquitectura de una Red Neuronal Convolutiva.

Con López Pacheco, M. A. (2017).

2.4.5. Fases de desarrollo Modelos de Machine Learning.

2.4.5.1. Fase de entrenamiento.

En esta fase se tiene una cantidad enorme de datos, de la cual se separa una parte para entrenar al algoritmo y darle toda esta información para que encuentre los patrones necesarios y después pueda hacer predicciones (Sandoval Serrano, L. J. 2018).

2.4.5.2. Fase de prueba.

El resto de los datos que quedan, se van a usar para hacer las pruebas. Así le podemos hacer preguntas al algoritmo y evaluar si las respuestas están bien o mal, y saber si está aprendiendo o no. Si vemos que no coinciden los datos, tendremos que agregar más datos o cambiar el método que estamos utilizando. Pero si se observa que hay entre un 80% a 90% de respuestas correctas,

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 39 de 128

podemos decir que hay un buen grado de aprendizaje y poder utilizar este algoritmo (Sandoval Serrano, L. J. 2018).

2.4.6. Visión Artificial.

La visión artificial es un área interesante y en constante desarrollo en inteligencia artificial. Con el desarrollo adicional de la tecnología, las computadoras y los sistemas han adquirido la capacidad de "ver" e interpretar imágenes y videos que les permiten comprender el mundo visual de una manera increíble.

A continuación, se describe parte del concepto.

La visión artificial es un campo de la inteligencia artificial (IA) que permite a los ordenadores y sistemas extraer información significativa a partir de imágenes digitales, videos y otras entradas visuales, y tomar medidas o realizar recomendaciones en función de esa información. Si la IA permite a los ordenadores pensar, la visión artificial les permite ver, observar y comprender. (¿Qué es la visión artificial? IBM. 2023)

De acuerdo con la información ofrecida en la página oficial de Microsoft Azure, ¿Qué es la visión artificial? (2023) indica que “El amplio abanico de aplicaciones prácticas de la tecnología de visión artificial la convierte en un componente central de numerosas innovaciones y soluciones modernas. La visión artificial se puede usar en la nube o en el entorno local”.

- ¿Cómo funciona?

Las aplicaciones de visión artificial utilizan entradas que proceden de sensores, inteligencia artificial, aprendizaje automático y aprendizaje profundo para replicar el funcionamiento de la

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 40 de 128

visión humana. Las aplicaciones de visión artificial ejecutan algoritmos que se entrenan con enormes cantidades de datos visuales o imágenes en la nube. Reconocen patrones en los datos visuales y usan esos patrones para determinar el contenido de otras imágenes. (¿Qué es la visión artificial? | Microsoft Azure. 2023)

- Algunas aplicaciones

1. *Google translate* permite a los usuarios apuntar la cámara de un teléfono inteligente a un cartel en otro idioma y obtener una traducción del mismo a su idioma preferido casi al instante. (¿Qué es la visión artificial? IBM. 2023).
2. El desarrollo de vehículos autónomos se basa en visión artificial para dar sentido a la entrada visual de las cámaras y otros sensores de un automóvil. Es imprescindible identificar otros vehículos, señales de tráfico, marcas de carriles, peatones, bicicletas y demás información visual que se encuentre en la carretera. (¿Qué es la visión artificial? IBM. 2023).

2.4.7. Procesamiento digital de imágenes.

“Es el conjunto de técnicas que se aplican a las imágenes digitales con el objetivo de mejorar la calidad o facilitar la búsqueda de información” (“Procesamiento digital, - EcuRed”. 2023)

El procesamiento de imágenes digitales es el uso de una computadora digital para procesar imágenes digitales a través de un algoritmo. Como subcategoría o campo del procesamiento de señales digitales, el procesamiento de imágenes digitales tiene muchas ventajas sobre el

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 41 de 128

procesamiento de imágenes analógicas. Permite aplicar una gama mucho más amplia de algoritmos a los datos de entrada y puede evitar problemas como la acumulación de ruido y distorsión durante el procesamiento. (“Procesamiento, AcademiaLab” 2023)

2.4.7.1. Pasos para el procesamiento de imágenes.

1. Importación de la imagen a través de herramientas de adquisición de imágenes (“Procesamiento digital, - EcuRed”. 2023)
2. Analizar y manipular la imagen (“Procesamiento digital, - EcuRed”. 2023)
3. Salida cuyo resultado puede ser una imagen alterada o un informe basado en el análisis de esa imagen (“Procesamiento digital, - EcuRed”. 2023).

2.4.7.2. Tipos de una imagen.

- Imagen binaria: la imagen binaria, como sugiere su nombre, contiene solo dos elementos de píxeles, es decir, 0 y 1, donde 0 se refiere al negro y 1 se refiere al blanco (“Conceptos básicos – Barcelona Geeks”. 2023)
- Imagen en blanco y negro: la imagen que consta solo de color blanco y negro (“Conceptos básicos – Barcelona Geeks”. 2023)
- Formato de color de 8 bits +: es el formato de imagen más famoso. Tiene 256 tonos diferentes de colores y se conoce comúnmente como imagen en escala de grises. En este formato, 0 representa negro, 255 representa blanco y 127 representa gris (“Conceptos básicos – Barcelona Geeks”. 2023)

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 42 de 128

- Formato de color de 16 bits es un formato de imagen en color. Tiene 65.536 colores diferentes (“Conceptos básicos – Barcelona Geeks”. 2023)

2.4.7.3. Ventajas.

- Método simple y sencillo de implementar. (“Procesamiento digital, - EcuRed”. 2023)
- Fácil asociación del concepto de frecuencia con ciertas características de la imagen; cambios de tonalidad suaves implican frecuencias bajas y cambios bruscos frecuencias altas. (“Procesamiento digital, - EcuRed”. 2023)
- Proporciona flexibilidad en el diseño de soluciones de filtrado. (“Procesamiento digital, - EcuRed”. 2023)
- Rapidez en el filtrado al utilizar el Teorema de la Convolución. (“Procesamiento digital, - EcuRed”. 2023)

2.5. Metodologías Machine learning

En el campo del proceso de aprendizaje automatizado, aprendizaje por medio de algoritmos de entrenamiento, patrones o en general de machine learning se pueden encontrar diferentes metodologías. A continuación, se describirán cinco de ellas:

2.5.1. CRISP-DM (Cross Industry Standard Process for Data Mining)

Según *International Business Machines* (IBM Documentation, 2021) CRISP-DM “es un método probado para orientar sus trabajos de minería de datos”. En otras palabras, es una metodología flexible y escalable que describe las fases de un proyecto y las tareas en cada una de ellas. Esta puede ser utilizada en proyectos de diferentes proyecciones y tamaños.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 43 de 128

El modelo de CRISP-DM es flexible y se puede personalizar fácilmente. Por ejemplo, si su organización intenta detectar actividades de blanqueo de dinero, es probable que necesite realizar una criba de grandes cantidades de datos sin un objetivo de modelado específico. En lugar de realizar el modelado, su trabajo se centrará en explorar y visualizar datos para descubrir patrones sospechosos en datos financieros. CRISP-DM permite crear un modelo de minería de datos que se adapte a sus necesidades concretas. (IBM Documentation, 2021)

De acuerdo con Chapman, P. & Clinton, J (2000)., el ciclo vital del modelo contiene seis etapas principales:

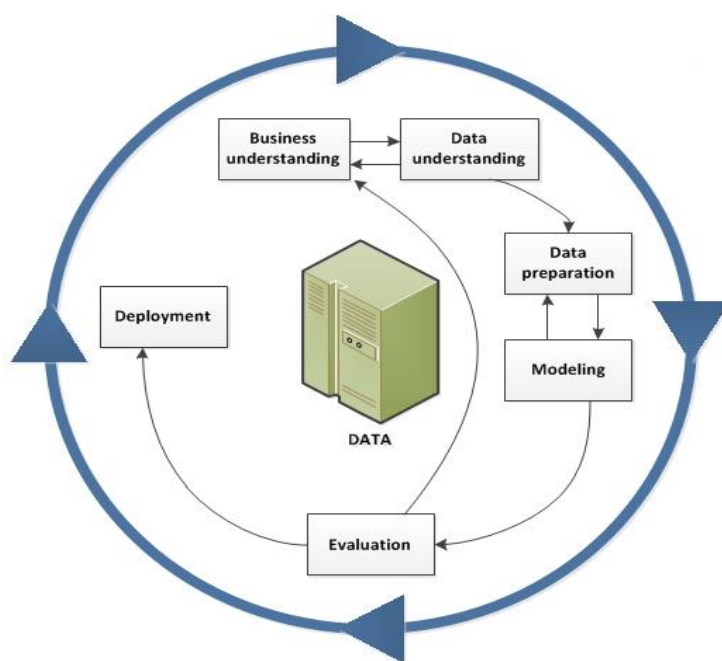


Imagen 7. Ciclo de vida de minería de datos.

IBM Documentation. (2023, mayo 17)

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 44 de 128

2.5.1.1. Comprensión del problema o negocio.

De acuerdo con IBM Documentation (2021), en esta etapa se debe “dedicar tiempo a explorar las expectativas de su organización con respecto a la minería de datos”, determinar los objetivos comerciales, evaluar la situación, determinar los objetivos de minería de datos, producir un plan de proyecto. “Las cuestiones que haya planteado hasta el momento y los objetivos comerciales y de minería de datos que haya formulado formarán la base de este plan” (IBM Documentation, 2021)

2.5.1.2. Comprensión de datos.

Se concluye lo siguiente: En esta fase se lleva a cabo la recolección y exploración inicial de los datos, con el objetivo de establecer un primer contacto con el problema. Esta fase suele ser crítica en el proyecto dado que un mal entendimiento de los datos tiene como consecuencia un aumento en el tiempo global del proyecto y también reduce las garantías de éxito (Daniel Alvarez, 2021)

2.5.1.3. Preparación de datos.

Juan Francisco Vallalta en su artículo (Juan Francisco Vallalta, 2019) publicado en la escuela de formación en inteligencia artificial en salud menciona que en esta fase se deberían obtener los datos finales, confirmando que:

En esta fase deberemos ser capaces de:

- Establecer el universo de datos con los que trabajar.
- Realizar tareas de limpieza de datos.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 45 de 128

- Construir un juego de datos apto para ser usado en modelos de minería de datos.
- Integrar datos de fuentes heterogéneas si es necesario.

2.5.1.4. Modelado.

En esta fase se plantea alcanzar los objetivos planteados para el proyecto. IBM Documentation (2021) propone el siguiente planteamiento en esta fase.

El modelado se suele ejecutar en múltiples iteraciones. Normalmente, los analistas de datos ejecutan varios modelos utilizando los parámetros predeterminados y ajustan los parámetros o vuelven a la fase de preparación de datos para las manipulaciones necesarias por su modelo. Es extraño que las cuestiones relativas a la minería de datos de una empresa se solucionen satisfactoriamente con un modelo y ejecución únicos.

2.5.1.5. Evaluación del modelo.

En esta fase se deben evaluar, para ello se debe tener en cuenta los “resultados de sus esfuerzos utilizando los criterios de rendimiento comercial establecidos en el inicio del proyecto” (IBM Documentation, 2021).

Según (Juan Francisco Vallalta, 2019) en esta fase se debe mínimamente poder realizar las siguientes partes:

- Evaluar el modelo o modelos generados hasta el momento.
- Revisar todo el proceso de minería de datos que nos ha llevado hasta este punto.
- Establecer los siguientes pasos a tomar, tanto si se trata de repetir fases anteriores como si se trata de abrir nuevas líneas de investigación.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 46 de 128

2.5.1.6. Implementación del modelo.

El objetivo último de esta fase es realizar el despliegue de los resultados obtenidos de forma que sea propagado a los usuarios finales, así como el mantenimiento de este una vez el despliegue haya finalizado (Juan Francisco Vallalta, 2019).

2.5.2. Proceso de ciencia de datos en equipo (TDSP)

El Proceso de Ciencia de Datos en Equipo (TDSP) es una metodología dinámica y repetitiva de ciencia de datos diseñada para ofrecer soluciones eficientes en análisis predictivo y aplicaciones inteligentes. TDSP promueve la colaboración y el aprendizaje en equipo al proporcionar directrices sobre cómo los diferentes roles dentro del equipo pueden trabajar de manera más efectiva en conjunto (Marktab, 2023).

Esta metodología incorpora prácticas y estructuras recomendadas por Microsoft y otros líderes de la industria, con el objetivo de garantizar la correcta implementación de iniciativas relacionadas con la ciencia de datos. En última instancia, TDSP busca ayudar a las organizaciones a reconocer y aprovechar las ventajas de su programa de análisis (Marktab, 2023).

De acuerdo con la página oficial de Microsoft la metodología TDSP se enfoca en la colaboración y comunicación efectiva entre los miembros del equipo de ciencia de datos y con los interesados del proyecto. También se enfoca en la documentación adecuada y la gestión de cambios para garantizar que el proyecto sea escalable y sostenible a largo plazo.

El proceso de ciencia de datos en equipo (TDSP) es una metodología de ciencia de datos ágil e iterativa para proporcionar soluciones de análisis predictivo y aplicaciones

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 47 de 128

inteligentes de manera eficiente. TDSP ayuda a mejorar la colaboración y el aprendizaje en equipo al sugerir cómo los roles de equipo funcionan mejor juntos. (Marktab, 2023)

2.5.2.1. Principales componentes del TDSP

De acuerdo con la documentación oficial en *Azure Architecture Center*, la TDSP tiene los siguientes componentes principales:

1. Una definición de ciclo de vida de ciencia de datos
2. Una estructura de proyecto estandarizada
3. Infraestructura y recursos recomendados para proyectos de ciencia de datos
4. Herramientas y utilidades recomendadas para la ejecución de proyectos

(Marktab, 2023)

2.5.2.1.1. Una definición de ciclo de vida de la ciencia de datos.

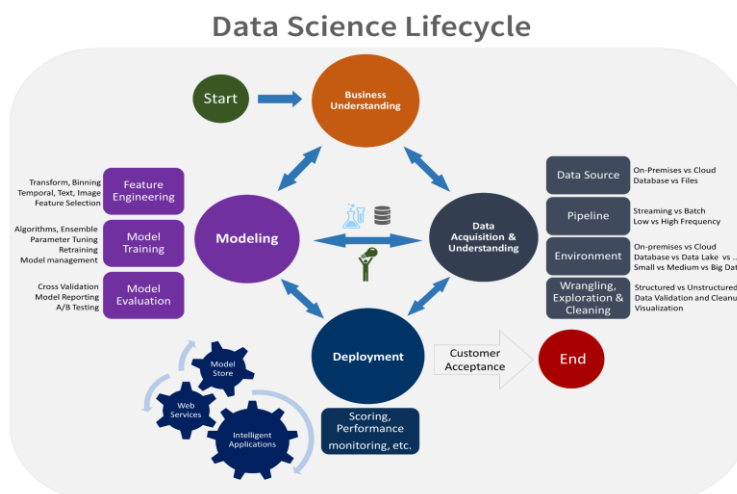


Imagen 8. Data Science Lifecycle.

Marktab. (2023).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 48 de 128

El TDSP, o proceso de ciencia de datos en equipo, ofrece una metodología que guía el desarrollo de proyectos de ciencia de datos. En este enfoque, se define un ciclo de vida que abarca todos los pasos necesarios para llevar a cabo proyectos de manera efectiva. (Fois, G., Agüero Crovella, G. A., y Britos, P. V. 2020).

2.5.2.1.2. Una estructura de proyecto estandarizada.

Tener una estructura de directorio y unas plantillas de proyecto definido ya estandarizadas son importantes para mantener la organización en el equipo que realice los proyectos de ciencia de datos. Esto va a permitir tener un acceso más fácil a toda la información, ya que por ejemplo al tener una misma estructura pueden acceder de manera rápida, por separado y confiable, con herramientas como Git o subversión (Marktab, 2023)

El tener las plantillas y la estructura de directorios ya estándar, permite a los miembros de la empresa comprender el trabajo de los demás y facilita a las personas nuevas un acoplamiento más sencillo (Marktab, 2023)

2.5.2.1.3. Infraestructura y recursos recomendados para proyectos de ciencia de datos.

TDSP proporciona recomendaciones para administrar análisis compartido e infraestructura de almacenamiento, por ejemplo:

- Sistemas de archivos en la nube para almacenar conjuntos de datos
- Data bases
- Clústeres de macrodatos (*SQL* o *Spark*)
- Servicio de aprendizaje automático

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 49 de 128

El uso de una infraestructura compartida evita la duplicación y costos innecesarios (Marktab. 2023).

2.5.2.1.4. Herramientas y utilidades recomendadas para la ejecución de proyectos.

Las herramientas proporcionadas para implementar el proceso y el ciclo de vida de la ciencia de datos ayudan a reducir las barreras a su adopción y la normalizan. Se ofrece un conjunto inicial de herramientas y scripts para automatizar algunas tareas comunes y se proporciona una estructura para que cada individuo aporte al repositorio compartido (Marktab. 2023)

2.5.3. KDD (Knowledge Discovery in Databases).

Es un proceso automático basado en los requerimientos del **usuario final**, en el que se combinan descubrimiento y análisis, con el objeto de extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice. (Flores Velásquez, D. A. 2019).

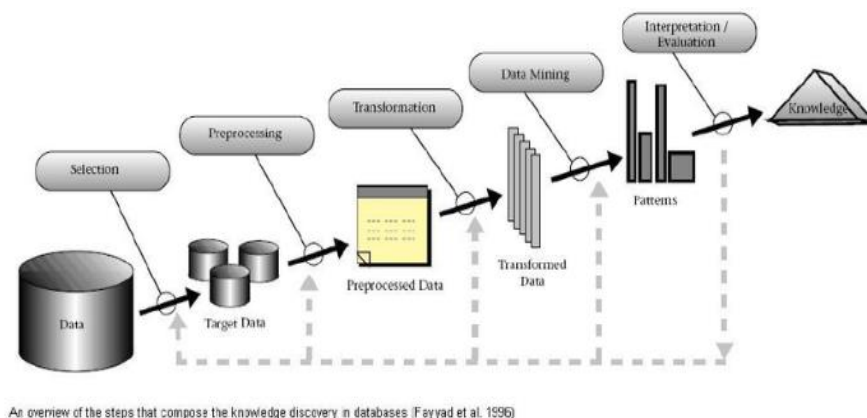


Imagen 9. The Knowledge Discovery in Databases (KDD) Process.

Fayyad (2019)

Se plantean las siguientes fases:

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 50 de 128

2.5.3.1. Integración y recopilación de datos.

En esta etapa se determinan las fuentes de información útiles y dónde conseguirlas. Las fuentes de información pueden residir en distintas bases de datos y pueden ser tanto internas como externas. También en esta fase se analizan los datos con el fin de darles un formato común, frecuentemente mediante el uso de un almacén de datos que permita unificar de manera operativa todos los datos, resolviendo la inconsistencia y facilitando su navegación, de esta manera será menos complejo comprender que aspectos resultan interesantes para el estudio. (Flores Velásquez, D. A. 2019)

2.5.3.2. Transformación/Reducción.

En esta etapa se seleccionan los datos a los que se le aplicarán los algoritmos de minería de datos, es crucial la cuidadosa selección de los atributos más importantes, ya que se eliminan y corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos. Por otro lado, en la limpieza se revisan los datos que no se ajustan al comportamiento general (*outliers*) y se decide cómo trabajarlos, también se eliminan los registros incoherentes y duplicados y se decide que tratamiento seguir con los faltantes o perdidos. Finalmente, en la transformación se construyen nuevos atributos con base en los ya existentes con el fin de mejorar el poder predictivo. (Flores Velásquez, D. A. 2019)

2.5.3.3. Minería de datos (Datamining).

Esta etapa del proceso KDD considera la aplicación de análisis de datos a los seleccionados en el paso anterior, el objetivo es descubrir patrones, tendencias y/o secuencias en los datos que

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 51 de 128

permitan extraer conocimiento de valor para el estudio. Por lo tanto, es aquí donde se decide sobre el tipo de tarea de minería de datos, el tipo de modelo y los algoritmos que serán utilizados para el descubrimiento y mayor comprensión de datos sin perder de vista el objetivo principal del estudio (Flores Velásquez, D. A. 2019).

2.5.3.4. Evaluación e interpretación.

Esta fase involucra todas las medidas de evaluación de resultados técnicos utilizando técnicas estadísticas como validación simple. Validación cruzada y *bootstrapping*. Por otra parte, también es necesario evaluar los resultados en torno al contexto en que serán utilizados, ya que no basta con la precisión de los algoritmos, contrastándolo con el conocimiento de expertos del negocio y si es necesario volver a etapas anteriores del proceso (Flores Velásquez, D. A. 2019).

2.5.3.5. Difusión y uso.

Finalmente, una vez validados los modelos se incorpora el nuevo conocimiento y se comunica a todos los usuarios relacionados al contexto del proyecto, con el objetivo de que el nuevo conocimiento se haga parte de la organización para su provecho (Flores Velásquez, D. A. 2019).

2.5.4. OSEM.

Es una estrategia comúnmente utilizada en ciencia de datos para resolver problemas de negocio y extraer información útil a partir de conjuntos de datos. Las fases de la metodología OSEMN son las siguientes (Han, 2019):

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 52 de 128

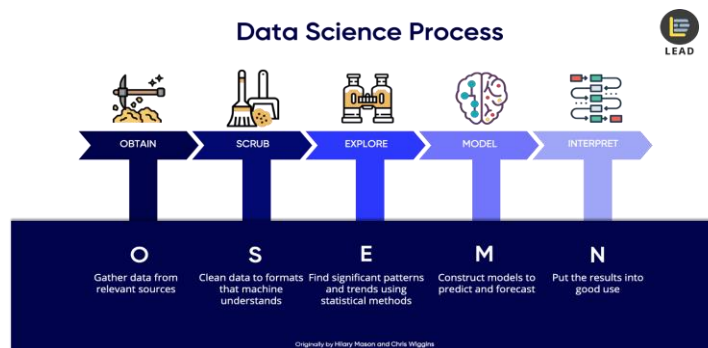


Imagen 10. Data Science Project.

Han, C. (2019, febrero 3).

2.5.4.1. Obtener los datos (O)

En esta fase se recopilan los datos necesarios para el análisis, ya sea a través de fuentes internas o externas. Es importante asegurarse de que los datos sean relevantes, precisos y estén en un formato utilizable (Han, 2019)

2.5.4.2. Limpiar y preparar los datos (S)

En esta fase se realiza una limpieza y preparación de los datos para asegurarse de que estén completos, consistentes y libres de errores. Esto incluye la eliminación de valores atípicos, la corrección de errores de sintaxis y ortográficos, y la normalización de los datos (Han, 2019).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 53 de 128

2.5.4.3. Explorar los datos (E)

En esta fase se realizan análisis exploratorios para comprender mejor los datos y encontrar patrones y tendencias. Esto incluye la visualización de datos, la identificación de relaciones y la exploración de características importantes (Han, 2019).

2.5.4.4. Modelar los datos (M)

En esta fase se utilizan técnicas de modelado estadístico y de aprendizaje automático para construir modelos que expliquen los patrones encontrados en los datos. Esto incluye la selección de variables importantes, la construcción de modelos predictivos y la validación de modelos (Han, 2019).

2.5.4.5. Interpretar los resultados (N)

En esta fase se interpretan los resultados del análisis para comprender los patrones y tendencias encontrados en los datos. Esto incluye la identificación de posibles causas y efectos, la evaluación de los modelos construidos y la toma de decisiones informadas basadas en los resultados (Han, 2019).

2.5.5. La metodología Canvas

Es una herramienta visual que permite identificar y definir los elementos clave de un modelo de negocio (Guerrero Paredes, M. D. P. 2017). Estos son los pasos que se suelen seguir en la metodología Canvas:

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 54 de 128

1. Segmentos de clientes: Identificación de los diferentes grupos de clientes a los que se quiere llegar con el negocio (Guerrero Paredes, M. D. P. 2017).

2. Propuesta de valor: Descripción detallada de los productos o servicios que se ofrecen a los clientes, y cómo se diferencian de la competencia (Guerrero Paredes, M. D. P. 2017).

3. Canales: Definición de los medios a través de los cuales se llegará a los clientes, como redes sociales, correo electrónico, tiendas físicas, etc. (Guerrero Paredes, M. D. P. 2017).

4. Relación con el cliente: Establecimiento de la forma en la que se relaciona el negocio con los clientes, como por ejemplo a través de atención al cliente, programas de fidelización, etc. (Guerrero Paredes, M. D. P. 2017).

5. Fuentes de ingresos: Identificación de las diferentes fuentes de ingresos del negocio, como la venta de productos o servicios, publicidad, comisiones, etc. (Guerrero Paredes, M. D. P. 2017).

6. Recursos clave: Identificación de los recursos clave necesarios para el funcionamiento del negocio, como recursos humanos, tecnológicos, financieros, etc. (Guerrero Paredes, M. D. P. 2017).

7. Actividades clave: Identificación de las actividades principales que se realizan para llevar a cabo el negocio, como la producción, la distribución, el marketing, etc. (Guerrero Paredes, M. D. P. 2017).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 55 de 128

8. Asociaciones clave: Identificación de las alianzas o asociaciones que pueden ser beneficiosas para el negocio, como proveedores, distribuidores, etc (Guerrero Paredes, M. D. P. 2017).

METODOLOGIA	FASES	ENFOQUE	VENTAJAS	DESVENTAJAS
CRISP-DM	CRISP-DM es un proceso de 6 pasos: Comprender la declaración del problema. Comprender los datos. Preparación de los datos. Realizar análisis de datos. Validación de los datos. Presentación/visualización de los datos.	Análisis de datos	El modelo de CRISP-DM es flexible y se pueden personalizar fácilmente. Como modelo de proceso, CRISP-DM ofrece un resumen del ciclo vital de minería de datos. (IBM Documentation. (2021)) Puede ser especialmente útil para planificar y explicar la gestión y ejecución de los proyectos a determinados clientes. (Haya, P. 2022)	Puede ser demasiado rígida y no adaptarse a todos los proyectos. Carece de procesos específicos para la gobernanza de los datos, tales como la recolección, el almacenamiento, versionado, la seguridad y la privacidad de los datos incluyendo el cumplimiento regulatorio y legal, que en el caso de datos personales ha cobrado una relevancia fundamental. (Haya, P. 2022)
TDSP	Presenta 4 fases: Una definición de ciclo de vida de ciencia de datos Una estructura de proyecto estandarizada Infraestructura y recursos recomendados para proyectos de ciencia de datos Herramientas y utilidades recomendadas para la ejecución de proyectos.	Trabajo en equipo	TDSP ayuda a mejorar la colaboración y el aprendizaje en equipo al sugerir cómo los roles de equipo funcionan mejor juntos. (“¿Qué es el Proceso,” 2021)	Desarrollada por Microsoft, lo que puede limitar su uso fuera de la plataforma Microsoft
AI CANVAS	Propone 8 elementos: Identificar los segmentos de clientes. Definir la propuesta de valor. Establecer los canales de distribución. Determinar la relación con los clientes. Especificar las fuentes de ingresos. Identificar los recursos clave. Establecer las actividades clave. Identificar las alianzas estratégicas. Evaluar la estructura de costos	Diseño de proyectos de AI	Permite identificar y definir los elementos clave de un modelo de negocio. Guerrero Paredes, M. D. P. (2017). El AI Project Canvas lo ayudará a responder las preguntas más apremiantes sobre el resultado y los recursos necesarios para su proyecto de IA. (Zawadzki, J. 2021) Proporciona una vista general y estructurada de un proyecto o negocio, lo que facilita la comunicación y colaboración entre los miembros del equipo. (Academia Win. 2023)	Al ser una herramienta visual, puede haber limitaciones para expresar ciertas ideas de manera detallada o rigurosa, lo que puede ser un inconveniente en proyectos que requieren un alto grado de especificidad y precisión. (Academia Win. 2023)

METOD OLOGÍA	FASES	ENFOQUE	VENTAJAS	DESVENTAJAS
KDD	Propone 5 fases: Integración y recopilación de datos. Transformación/ Reducción. Minería de datos(Dataminin g). Evaluación e interpretación. Difusión y uso.	Descubrimi ento de conocimient os	Alberga muchos enfoques diferentes para el descubrimiento, que incluye aprendizaje inductivo, estadística bayesiana, optimización de consultas semánticas, adquisición de conocimiento para sistemas expertos y teoría de la información. (Knowledge Discovery, . 2017)	De acuerdo con ChatGPT (2021) las principales desventajas son: Complejidad y tiempo de procesamiento: La metodología KDD puede ser compleja y requiere tiempo para completar todos los pasos, lo que puede ser desafiante y consumir recursos significativos. Riesgo de overfitting: Existe la posibilidad de que los modelos generados mediante KDD se ajusten demasiado a los datos de entrenamiento, lo que puede resultar en un rendimiento deficiente en nuevos datos. Dependencia de la calidad y disponibilidad de los datos: La calidad y disponibilidad de los datos utilizados en KDD son fundamentales para obtener resultados precisos y confiables. Los datos incompletos, ruidosos o sesgados pueden afectar.
OSEM	6 fases secuenciales. Obtener los datos (O) Limpiar y preparar los datos (S) Explorar los datos (E) Modelar los datos (M) Interpretar los resultados (N)	Herramient as de código abierto	Simple: destila el complejo proceso de un proyecto de ciencia de datos en cinco pasos claros. Esto es especialmente notable dado que este proceso general y el concepto moderno de ciencia de datos aún eran nuevos cuando Mason y Wiggins crearon OSEMN en 2010. Tiene sentido: los pasos presentados tienen un flujo lógico representativo del ciclo de vida general de la ciencia de datos.	Sin datos, no tiene un proyecto de análisis de datos o ciencia de datos. Pierde la comprensión del negocio: el marco comienza con Obtener, que ignora las preguntas básicas clave que deberían venir primero, a saber: "¿Debo invertir tiempo en este proyecto?" y "¿Qué resultado estoy tratando de impulsar?" No considera la implementación: OSEMN asume implícitamente que está entregando una salida única. En realidad, a menudo es necesario implementar un modelo en un sistema de producción para que continúe proporcionando valor a lo largo del tiempo. (Hotz, N. 2023)

Tabla 2. Cuadro comparativo metodologías Machine Learning.

2.6. Modelos de detección de objetos

La detección de objetos es una tarea en el campo de la visión por computadora que implica identificar y localizar objetos específicos en una imagen o un video. Consiste en entrenar modelos

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 57 de 128

de aprendizaje automático, como redes neuronales convolucionales, para reconocer y delimitar objetos de interés en una imagen (Villanueva, R. M. 2017)

La detección de objetos implica varios pasos, incluyendo la extracción de características de la imagen, la generación de propuestas de regiones de interés, el refinamiento de las detecciones y la clasificación de los objetos en diferentes categorías.

Los modelos de detección de objetos, como Yolo, SSD (*Single Shot Multibox Detector*) o Faster R-CNN (*Región Convolutacional Neural Network*), han logrado avances significativos en esta tarea, permitiendo la detección precisa y eficiente de objetos en tiempo real. (Villanueva, R. M. 2017)



Imagen 11. Modelo de detección de objetos en tiempo real.

Mario Massaro (2021, June 10).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 58 de 128

A continuación, se realizará una descripción de algunos algoritmos de detección de objetos, como SSD, *Faster R-CNN* y *Yolo*. Estos algoritmos son ampliamente utilizados en aplicaciones de detección de objetos debido a sus características y rendimiento. La selección de estos algoritmos se basa en su popularidad en la comunidad científica y su éxito en diversas tareas de detección de objetos.

Se analizará cada uno de ellos para determinar cuál sería la mejor opción para la detección de la Varroosis en las abejas. El objetivo es encontrar un algoritmo preciso y eficiente que pueda identificar de manera confiable las abejas infectadas con Varroosis y ayudar a proteger la salud de las colonias de abejas polinizadoras.

2.6.1. SSD - Single Shot Detector.

SSD se basa en una red CNN de propagación hacia adelante, que produce una serie de cuadros delimitadores de tamaño fijo, y la posibilidad de contener instancias de objetos en cada cuadro, es decir, puntuación. (Algoritmo de detección, 2023)

Tiene una estructura piramidal en su CNN en la que las capas van disminuyendo gradualmente. Esto le permite poder detectar objetos grandes y pequeños. No utiliza una grilla predefinida, pero cuenta con “anclas” de distintas proporciones que se van escalando a medida que descendemos por la pirámide (mapa de *features* más pequeños, con anclas proporcionalmente más grandes). (Modelos de Detección. 2020)

El modelo SSD se compone de 2 partes, a saber:

- El modelo de la columna vertebral.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 59 de 128

- El cabezal SSD.

El modelo *Backbone* es una red típica de clasificación de imágenes previamente entrenada que funciona como extractor de mapas de entidades. Aquí, las capas de clasificación de imágenes finales de imagen del modelo se eliminan para darnos solo los mapas de entidades extraídos.

El cabezal SSD se compone de un par de capas convolucionales apiladas juntas y se agrega a la parte superior del modelo de columna vertebral. Esto nos da la salida como los cuadros delimitadores sobre los objetos. Estas capas convolucionales detectan los diversos objetos de la imagen.

Ventajas

- SSD es muy simple de entrenar y se puede integrar fácilmente en sistemas que realizan la detección como una de sus funciones. (Srivastava, S., Divekar. 2021)
- El modelo SSD detecta objetos en una sola pasada, lo que significa que ahorra mucho tiempo. Pero al mismo tiempo, el modelo SSD también parece tener una precisión asombrosa en su detección. (Srivastava, S., Divekar. 2021)
- SSD se puede entrenar de extremo a extremo para una mejor precisión. (Sojasingarayar, A. 2023).
- SSD hace más predicciones y tiene una mejor cobertura en ubicación, escala y relaciones de aspecto.
- SSD funciona más rápido que los detectores en tiempo real (Yolo) de última generación y es tan preciso como Faster R-CNN. (SSD: Understanding. 2019)

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 60 de 128

Desventajas

- Otros modelos de detección de objetos como Yolo o Faster R-CNN realizan sus operaciones a una velocidad mucho menor en comparación con el SSD. (Srivastava, S., Divekar. 2021)
- SSD tiene sus limitaciones de detección de objetos pequeños debido a la pérdida de sus características semánticas, este inconveniente es relativamente remediado por una característica significativa implementada en el documento de RetinaNet que implementa un FPN (Feature Pyramid Network). (SSD: Understanding single, 2019).

2.6.2. *Faster R-CNN.*

R-CNN significa Redes neuronales convolucionales basadas en regiones. Este método combina propuestas de región para la segmentación de objetos y CNN de alta capacidad para la detección de objetos. (Reza, Z. N. 2019)

R-CNN más rápido es un algoritmo de detección de objetos propuesto por Kaiming Ren, Kaiming He, Ross Girshick y Jian Sun en 2015. (Detección de objetos, 2023)

Faster R-CNN es un único modelo unificado, la arquitectura se compone de dos módulos:

- RPN (Red de Propuestas de Región): Red neuronal Convolucional para proponer regiones y el tipo de objeto a considerar en la región. (Sojasingarayar, A. 2023)
- Rápido R-CNN: Red neuronal Convolucional para extraer características de las regiones propuestas y generar el cuadro delimitador y las etiquetas de clase. (Sojasingarayar, A. 2023).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 61 de 128

Ventajas

- La red puede predecir con precisión y rapidez las ubicaciones de diferentes objetos (Sojasingarayar, A. 2023).
- Fast R-CNN es más rápido que su predecesor porque no se requiere alimentar ~ 2000 propuestas como entrada a la CNN por ejecución. La operación de convolución se realiza para generar un mapa de entidades solo una vez por imagen. (Srivastava, S., Divekar, 2021)
- La red puede predecir con precisión y rapidez las ubicaciones de diferentes objetos (Sojasingarayar, A. 2023).

Desventajas

- Este algoritmo muestra una reducción significativa en el tiempo requerido tanto para el entrenamiento como para las pruebas en comparación con R-CNN. Pero se observó que incluir propuestas de región obstaculiza significativamente el algoritmo, reduciendo su rendimiento (Srivastava, S., Divekar, 2021)
- La red no es lo suficientemente rápida como para ser utilizada como un sistema en tiempo real porque tarda aproximadamente (2 segundos) en generar salida en una imagen de entrada (Faster R CNN ML. 2022).

2.6.3. Yolo

En los tiempos modernos, Yolo es uno de los algoritmos de detección de objetos más precisos y exactos disponibles (Srivastava, S., Divekar, 2021).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 62 de 128

En Yolo se toma la detección de objetos como un problema único de regresión, una única red Convolutiva predice simultáneamente múltiples cuadros delimitadores que enmarcan los objetos en la imagen y predice probabilidades condicionales por cada clase p (Clase | Objeto) para cada uno de estos cuadros delimitadores (Massiris, Delrieux y Fernández, 2020). La red neuronal puede lograr una velocidad de ejecución de 45 fotogramas por segundo (fps) en computadoras de propósito general (Massiris, Delrieux y Fernández, 2020).

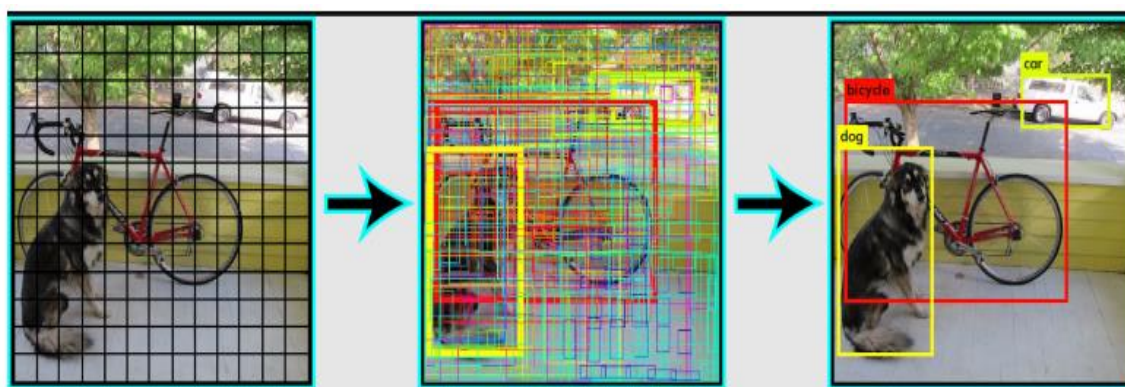


Imagen 12. Introducción a Yolo: detección de objetos en tiempo real.

HashDork Jay. 2022.

YoloV5 utiliza una red troncal de red neuronal Convolutiva (CNN) para formar características de imagen. Estas características se combinan en el cuello del modelo y se envían a la cabeza. A continuación, el cabezal del modelo interpreta las características combinadas para predecir la clase de una imagen. (Solawetz, J. 2023).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 63 de 128

2.6.3.1. Arquitectura del modelo Yolo.

La red de detección tiene 24 capas convolucionales seguidas de 2 capas completamente conectadas. La alternancia de capas convolucionales 1×1 reduce el espacio de características de las capas anteriores. Se pre entrenan las capas convolucionales en la tarea de clasificación de Imagen Neta la mitad de la resolución (imagen de entrada 224×224) y luego se duplica la resolución para la detección (“Vista de Introducción al Aprendizaje,” 2023).

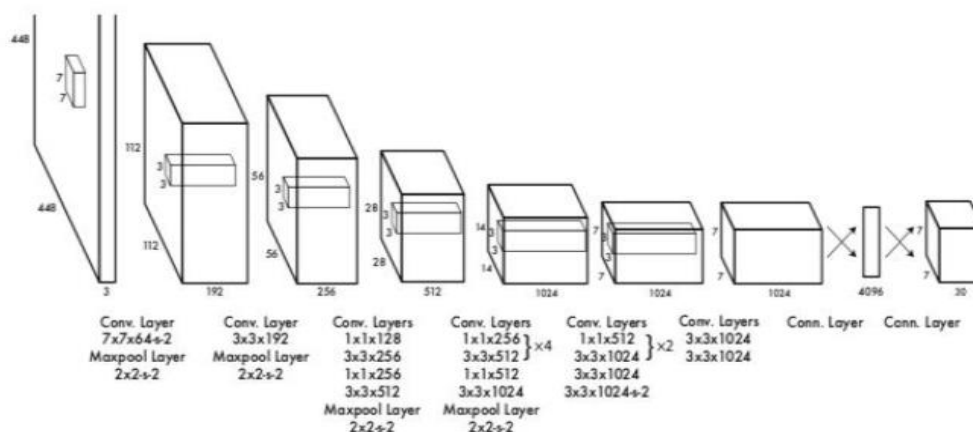


Imagen 13. Arquitectura del sistema Yolo.

(Yolo. 2023).

Yolo divide la imagen en una cuadrícula y realiza cálculos para obtener valores como las probabilidades de las clases y los parámetros de las cajas delimitadoras. El modelo funciona dividiendo la imagen de entrada en una cuadrícula de celdas, donde cada celda se encarga de predecir una caja delimitadora si el centro de dicha caja cae dentro de la celda. Cada celda de la cuadrícula predice una caja delimitadora que incluye las coordenadas x, y, el ancho, el alto y la

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 64 de 128

confianza. Además, se realiza una predicción de la clase correspondiente a cada celda (Sojasingarayar, A. 2023)

Ventajas

- Gracias a estos retoques, logra la sorprendente capacidad de casi 60 FPS (cuadros por segundo) en ordenadores normales (Modelos de Detección. 2020)
- YoloV5 es muy fácil de usar y viene listo para usar en objetos personalizados "listos para usar" (Solawetz, J. 2023)
- YoloV5 se puede ejecutar en cualquiera de los siguientes entornos verificados actualizados (con todas las dependencias) incluyendo *CUDA/CUDNN*, *Python* y *PyTorch* preinstalados (*Ultralytics*. 2023)

Desventajas

- Se le critica que, si bien es rápida, suele tener menor porcentaje de aciertos frente a las R-CNN. (Modelos de Detección. 2020)

ALGORITMO	YOLO(YOU ONLY LOOK ONCE)	FASTER R-CNN	SSD (SINGLE SHOT MULTIBOX DETECTOR)
Precisión	Buena	Muy buena	Buena
Velocidad	Alta	Media	Alta
Tamaño del modelo	Pequeño	Grande	Mediano
Detección de objetos pequeños	Bueno	Bueno	Regular
Detección de objetos grandes	Bueno	Bueno	Bueno
Eficiencia computacional	Alta	Media	Alta
Requisitos de entrenamiento	Menos datos y tiempo requeridos	Más datos y tiempo requeridos	Menos datos y tiempo requeridos
Arquitectura de red neuronal	Convolución completa	Basada en regiones	Convolución completa
Flexibilidad en la elección del tamaño de la entrada	Limitada	Flexible	Flexible

Tabla 3. Cuadro comparativo Modelos de detección de imágenes.

OpenAI. (2023).

2.6.3.2. Evolución del modelo Yolo.

2.6.3.2.1. Yolo Versión 1.

Fue la versión original de Yolo presentada en 2015. Introdujo el concepto de detección de objetos en una sola pasada de la red neuronal Convolucional y fue pionero en la detección de

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 66 de 128

objetos en tiempo real. Sin embargo, Yolo v1 tuvo problemas con la localización precisa de objetos pequeños y la detección de objetos superpuestos. (Redmon, Divvala, Girshick y Farhadi, 2016).

2.6.3.2.2. Yolo Versión 2.

Yolo (v2) es conocido como un sucesor inmediato de Yolo (v1), utilizado para detectar objetos en alrededor de 9000 categorías. La mayoría de las técnicas de detección de objetos existentes están restringidas a la clasificación de categorías limitadas. Esto se debe ciertamente a la escasez de datos etiquetados para la tarea de detección de objetos. Entonces, los autores probaron la escalabilidad en la tarea de detección de objetos para un mayor número de categorías (Diwan, G Anirudh y Tembhurne, 2022) Para hacerlo, se combinaron los conjuntos de datos de ImageN y COCO, lo que resultó en más de 9418 categorías de instancias de objetos.

El problema de los datos para la clasificación es enorme ahora, teniendo esto en cuenta, se desarrolla un método jerárquico conocido como árbol de palabras para combinar clasificación y detección. En esta sección, discutimos la arquitectura de Yolo (v2) y las mejoras de la versión base (Diwan, G Anirudh y Tembhurne, 2022).

La arquitectura Yolo (v2) está inspirada en VGG y Network-in-Network. Utiliza el marco darknet-19 que consta de 19 capas convolucionales y 5 capas de agrupación máximas, Además, también se observa muchas convoluciones 1×1 para lograr un muestreo descendente en toda la profundidad del volumen de entrada (Diwan, G Anirudh y Tembhurne, 2022). Tiene muchas características adicionales en comparación con la versión base. Varias técnicas de aumento de

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 67 de 128

datos, como cultivos aleatorios, rotaciones y muchas más, se emplean para el entrenamiento de modelos; sin embargo, esta versión tiene problemas para detectar objetos de menor tamaño.

Además de utilizar las funciones existentes, como la agrupación promedio global y la convolución 1×1 , los autores también introdujeron nuevas técnicas de optimización, como la normalización por lotes, el clasificador de alta resolución, la convolución con cuadros de anclaje, los clústeres de dimensiones, la predicción de ubicación directa, las funciones de grano fino y Multi El entrenamiento a escala, se puede describir brevemente en la siguiente imagen (Diwan, G Anirudh y Tembhrne, 2022)

Type	Filters	Size/Stride	Output
Convolutional	32	3×3	224×224
Maxpool		$2 \times 2/2$	112×112
Convolutional	64	3×3	112×112
Maxpool		$2 \times 2/2$	56×56
Convolutional	128	3×3	56×56
Convolutional	64	1×1	56×56
Convolutional	128	3×3	56×56
Maxpool		$2 \times 2/2$	28×28
Convolutional	256	3×3	28×28
Convolutional	128	1×1	28×28
Convolutional	256	3×3	28×28
Maxpool		$2 \times 2/2$	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Maxpool		$2 \times 2/2$	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	1000	1×1	7×7
Avgpool		Global	1000
Softmax			

Imagen 14. Operaciones arquitectónicas por capas en el marco Darknet-19.

Diwan, T., G Anirudh y Tembhrne, J. V. (2022).

2.6.3.2.3. Yolo Versión 3.

La versión base de Yolo no tenía una solución para los errores de localización y la segunda versión no pudo detectar los objetos de menor tamaño (Diwan, G Anirudh y Tembhrne, 2022).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 68 de 128

La tercera versión de Yolo intenta superar los inconvenientes antes mencionados y proporciona una forma eficiente de detectar objetos con un rendimiento mejorado mientras se entrena y evalúa en el conjunto de datos COCO. Sin embargo, como esta versión tiene un rendimiento superior para objetos de tamaño más pequeño, sufre al producir resultados precisos para objetos de tamaño mediano y grande (Diwan, G Anirudh y Tembhrne, 2022).

Esta arquitectura se basa en el marco Darknet-53, una red que utiliza un filtro Convolutacional de 3×3 y 1×1 junto con algunas conexiones de acceso directo, una arquitectura significativamente más grande y poderosa con 53 capas convolucionales. Es dos veces más rápido que ReNet-152 sin comprometer el rendimiento. La arquitectura genérica subyacente para Yolo (v3) (Diwan, G Anirudh y Tembhrne, 2022).

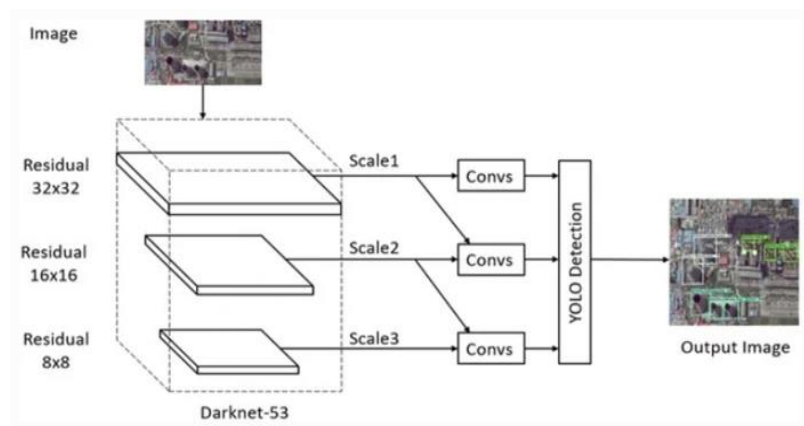


Imagen 15. Arquitectura de Yolo (v3).

Diwan, T., G Anirudh y Tembhrne, J. V. (2022).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 69 de 128

2.6.3.2.4. Yolo Versión 4.

Lanzado en 2020, Yolo v4 fue una versión altamente optimizada y de alto rendimiento. Utilizó una arquitectura denominada CSP Darknet 53, que combina capas convolucionales y conexiones en bloque, lo que mejoró la capacidad de representación de características. También introdujo atención espacial y temporal para capturar relaciones a largo plazo en videos y utilizó técnicas avanzadas de aumento de datos como Mosaic y Cut Mix (Bochkovskiy, Wang y Liao, 2020).

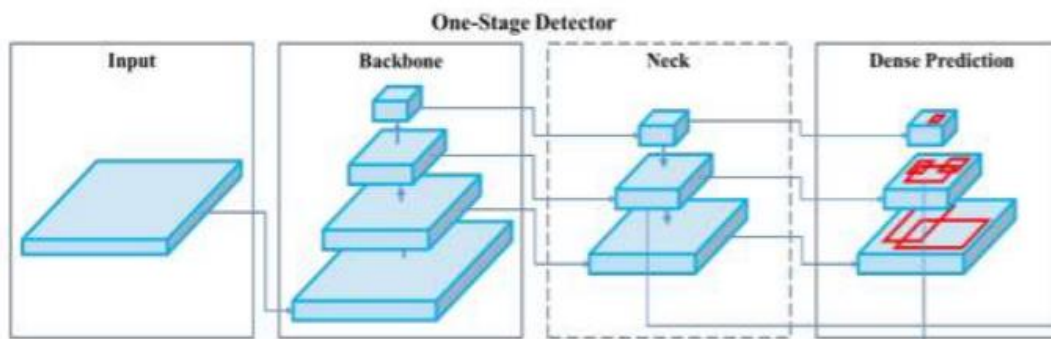


Imagen 16. Arquitectura de la red YoloV4.

Rozada Raneros, S. (2021).

En esta versión, su motivo principal es optimizar el detector para realizar cálculos en paralelo. se compone en tres etapas:

- *Backbone*: CSPDarknet53, una red que aumenta la capacidad de aprendizaje y que gracias a la unión del módulo de agrupación de pirámide espacial permite mejorar el campo receptivo y distinguir características importantes. Rozada Raneros, S. (2021).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 70 de 128

- *Neck*: Módulo de agrupación de pirámides espaciales y agregación de ruta PANet, principalmente. PANet se implementa como sustitución de las redes piramidales de características empleadas para la detección en Yolo v3. Rozada Raneros, S. (2021).
- *Head*: Yolo v3 Trabajo de Fin de Máster También encontramos dos conceptos que son: Bag of Specials y Bag of Freebees. Bag of specials se refiere a las mejoras que se realizan en el proceso de capacitación como la inferencia de datos, función de coste, pero que no presentan un impacto en la velocidad de inferencia. Mientras que Bag of specials se refiere al impacto marginal que tienen en el tiempo de inferencia y el buen rendimiento de las mejoras anteriores. Además, esto mejora el aprendizaje de las características. Rozada Raneros, S. (2021).

2.6.3.2.5. Yolo Versión 5.

También lanzado en 2020, YoloV5 fue una implementación más ligera y rápida de Yolo. Se centró en mejorar la velocidad de detección mientras mantenía un rendimiento competitivo. YoloV5 utilizó una arquitectura de detección de objetos de una sola etapa, pero con modificaciones y mejoras significativas. Fue desarrollado por *Ultralytics* y se ha convertido en una versión popular utilizada en aplicaciones de detección de objetos en tiempo real. (Ultralytics, 2023).

Es una familia de arquitecturas basado en la data sets COCO y escrito en Python. el cual usa esta la librería *PyTorch*. Además, presenta una gran variedad de modelos para ajustarse al uso que se vaya a dar, permitiendo el uso de algunos en pequeños dispositivos para la detección de objetos en tiempo real.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 71 de 128

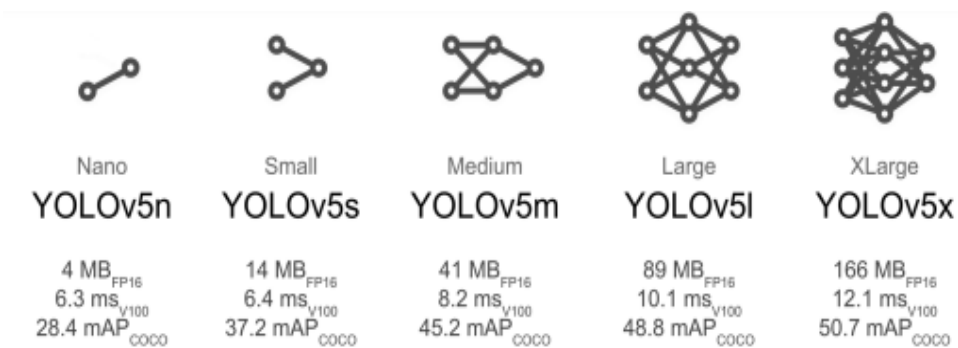


Imagen 17. Descripción de los modelos YoloV5.

De Jeu Boronat, C. (2022).

- YoloV5n: Es el modelo más pequeño de la familia, dedicado a dispositivos IoT, y dispositivos móviles con bajas prestaciones. García-García, P. (2022).
- YoloV5s: Un modelo pequeño pero que puede utilizarse en tiempo real. García-García, P. (2022).
- YoloV5m: Es la versión intermedia, ofreciendo un buen balance entre velocidad y precisión. YoloV5l: Un modelo grande e ideal para detectar objetos pequeños. García-García, P. (2022).
- YoloV5x: Es el modelo más grande de la familia, y ofrece la mejor precisión por encima del resto a cambio de ser el más lento. García-García, P. (2022).

La arquitectura YoloV5x, que es una poderosa red neuronal Convolutacional (CNN). Estas redes utilizan capas convolucionales para extraer características relevantes de las imágenes y realizar operaciones de convolución que les permiten identificar patrones y objetos en diferentes escalas (Massiris, Delrieux y Fernández, 2020). Es importante recalcar que el proyecto ha sido

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 72 de 128

desarrollado mediante el enfoque de transferencia de aprendizaje, la cual permite aprovechar un modelo pre entrenado en un conjunto de datos amplio y diverso para la detección de objetos en general. (Karl Heinz Weiss, Khoshgoftaar y Wang, 2016). A través de un proceso de *fine-tuning*, el modelo pre entrenado se ha adaptado específicamente a la tarea de detección de Varroosis en las abejas.

YoloV5x es una variante de la arquitectura Yolo que ha demostrado excelentes resultados en la detección de objetos en imágenes con una precisión y velocidad superiores (Wang et al., 2022). Se eligió YoloV5x debido a sus características avanzadas, como la capacidad de detectar múltiples objetos simultáneamente y su alta eficiencia computacional, lo que permite realizar detecciones en tiempo real. Además, YoloV5x ha sido entrenado en conjuntos de datos masivos y diversificados, lo que mejora su capacidad para reconocer y clasificar abejas infectadas con Varroosis de manera precisa y confiable.

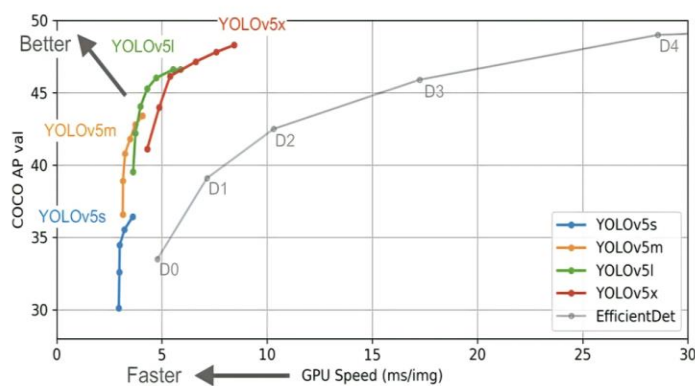


Imagen 18. YoloV5-P5 640.

Ultralytics. (2023, May 25)

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 73 de 128

En el gráfico, el objetivo es crear un modelo de detección de objetos altamente eficiente (eje Y) en relación con su tiempo de inferencia (eje X). Los resultados preliminares demuestran que YoloV5 tiene un desempeño excepcional en comparación con otras técnicas de vanguardia. En el gráfico, se puede observar que todas las variantes de YoloV5 se entrenan más rápido que *EfficientDet*. El modelo más preciso de YoloV5, llamado YoloV5x, puede procesar imágenes múltiples veces más rápido con una precisión similar al modelo *EfficientDet* (Solawetz, J. 2023).

2.6.3.2.6. Yolo Versión 6.

Yolo v6 fue publicado en ArXiv en septiembre de 2022 por el Departamento de IA de Meituan Visión. El diseño de la red consta de un *backbone* eficiente con bloques *RepVGG* o *CSPStackRep*, un cuello de topología PAN y un eficiente cabeza desacoplado con una estrategia de canal híbrido. Además, el documento presenta técnicas de cuantificación mejoradas, utilizando la cuantificación posterior al entrenamiento y la destilación por canales, lo que da como resultado detectores más rápidos y precisos (Terven & Cordova-Esparaza, 2023).

Como cualquier otro modelo de Yolo, el Yolo V6 también tiene tres componentes. Son la columna vertebral, el cuello y la cabeza de la red, y todos tienen algo nuevo que ofrecer. Uno de los aspectos más importantes de Yolo v6 es que no tiene anclaje y utiliza una columna vertebral re parametrizada. (Terven & Cordova-Esparaza, 2023).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 74 de 128

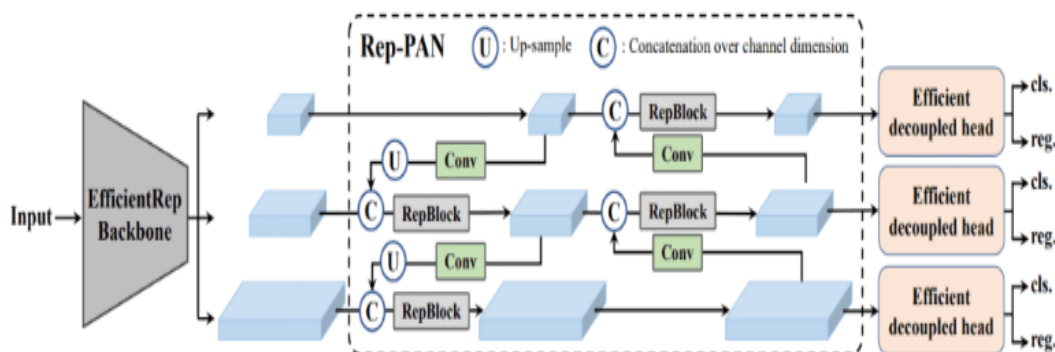


Imagen 19. Arquitectura del modelo YoloV6.

“Yolov6 Object Detection – Paper Explanation and Inference,” 2022.

2.6.3.2.7. Yolo Versión 7.

Fue lanzado en julio de 2022 por WongKinYiu y AlexeyAB. Logra un rendimiento de vanguardia y está capacitado para detectar las 80 clases genéricas en el conjunto de datos MS COCO para la detección de objetos en tiempo real (“YoloV5,” 2022). Además, tiene aplicaciones en la segmentación de casos y la evaluación de la postura humana. En la rama de la máscara, la segmentación de instancias de Yolo v7 es un enfoque *Single Stage*, lo que contribuye a su eficiencia en esta área. El *OrienMaskHead* se utiliza en este momento y es posible que se implementen más enfoques en el futuro (Cao, Zheng y Fang, 2023)

A continuación, se muestra el procedimiento de red del marco Yolo v7 en su totalidad. La imagen de entrada se redimensiona primero a 640×640 píxeles antes de alimentar la red troncal.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 75 de 128

A continuación, se generan tres capas de mapas de características de diferentes tamaños a través de la red de capa principal, y se utilizan Rep y conv para generar los resultados de la predicción. (Cao, Zheng y Fang, 2023). Si el conjunto de datos pertenece a Coco, cada salida (x, y, w, h, o) es la ubicación de las coordenadas y el telón de fondo antes y después, y hay tres anclas. Las salidas se dividen en 80 categorías. Por lo tanto, el resultado final de cada capa es $(80 + 5) \times 3 = 255$ multiplicado por el tamaño del mapa de características (Cao, Zheng y Fang, 2023).

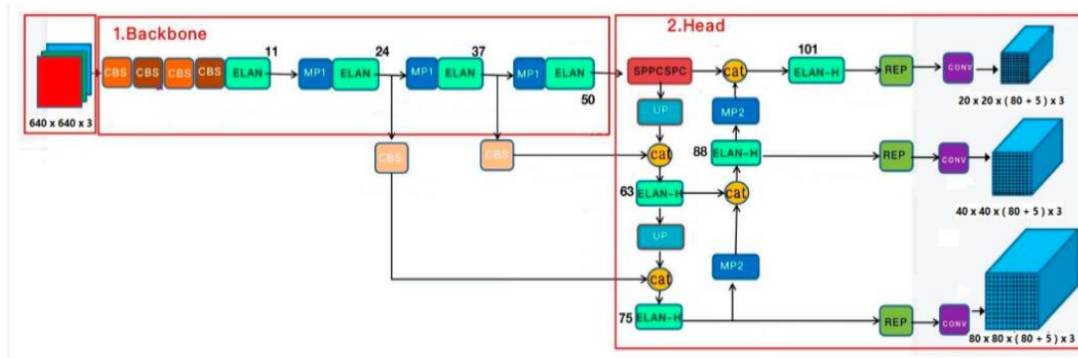


Imagen 20. Proceso de red de marco general de Yolo v7.

Cao, L., Zheng, X. y Fang, L. (2023).

2.6.3.2.8. Yolo Versión 8.

Fue lanzado en enero de 2023 por *Ultralytics*, la empresa que desarrolló YoloV5. Yolo v8 proporcionado cinco versiones escaladas: Yolo v8n (nano), Yolo v8s (pequeño), Yolo v8m (mediano), Yolo v8l (grande) y Yolo v8x (extra grande).

Yolo v8 admite múltiples tareas de visión, como detección de objetos, segmentación, estimación de poses, seguimiento, y clasificación.

El modelo usa una columna vertebral similar a la de YoloV5 con algunos cambios en el *CSPLayer*, ahora llamado módulo C2f. El módulo C2f (cuello de botella parcial entre etapas con dos circunvoluciones) combina características de alto nivel con información contextual para mejorar la precisión de detección. Utiliza un modelo sin anclaje con un cabezal desacoplado para procesar de forma independiente la objetualizada, la clasificación y tareas de regresión (Terven y

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 77 de 128

Cordova-Esparaza, 2023). Este diseño permite que cada rama se concentre en su tarea y mejora la precisión general del modelo.

En la capa de salida de Yolo v8, utilizaron la función sigmoidea como función de activación para la puntuación de objetividad, lo que representa la probabilidad de que el cuadro delimitador contenga un objeto. (Terven y Cordova-Esparaza, 2023). Utiliza la función *softmax* para las probabilidades de clase, representando las probabilidades de los objetos pertenecientes a cada clase posible. Yolo v8 utiliza las funciones de pérdida *CIoU* para la pérdida del cuadro delimitador y la entropía cruzada binaria para la clasificación pérdida. Estas pérdidas han mejorado el rendimiento de la detección de objetos, especialmente cuando se trata de objetos más pequeños. El modelo tiene cabezales de detección similares a Yolo v8, que consta de cinco módulos de detección y una capa de predicción (Terven y Cordova-Esparaza, 2023).

2.7. La transferencia de aprendizaje IA

Es un enfoque que busca aplicar el conocimiento y las habilidades adquiridas por un modelo pre entrenado en una tarea específica a una tarea o dominio diferente. Esta técnica se basa en la idea de que muchos aspectos del aprendizaje automático son generales y pueden aplicarse a una variedad de problemas. Al utilizar un modelo pre entrenado como punto de partida, se puede aprovechar el conocimiento previo que el modelo ha adquirido al reconocer patrones y estructuras en grandes conjuntos de datos (Karl Heinz Weiss, Khoshgoftaar y Wang, 2016).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 78 de 128

Este proceso implica ajustar el modelo mediante un proceso llamado "ajuste fino" o "reentrenamiento". Durante este proceso, el modelo se entrena adicionalmente en datos específicos de la nueva tarea o dominio para adaptarlo a esa tarea en particular. Esto permite que el modelo aproveche su conocimiento previo y acelere el proceso de entrenamiento en la nueva tarea (Karl Heinz Weiss, Khoshgoftaar y Wang, 2016). Este proceso es útil cuando se dispone de conjuntos de datos limitados o cuando entrenar un modelo desde cero sería costoso en términos de tiempo y recursos computacionales. Al aprovechar el conocimiento pre entrenado, se puede lograr un rendimiento superior con menos datos de entrenamiento (Karl Heinz Weiss, Khoshgoftaar y Wang, 2016).

Sin embargo, es importante tener en cuenta que la transferencia de IA no siempre es aplicable en todos los casos. Algunas tareas pueden requerir un ajuste más profundo o una reestructuración del modelo para adaptarse adecuadamente a la nueva tarea. Además, la transferencia de IA también puede presentar desafíos éticos y de privacidad, especialmente cuando se utilizan modelos entrenados en datos sensibles o en dominios donde se requiere un alto nivel de confianza y seguridad (Karl Heinz Weiss, Khoshgoftaar y Wang, 2016).

2.8. Google colab

Colab, también conocido como "*Colaboratory*", te permite programar y ejecutar Python en tu navegador con las siguientes ventajas:

- No requiere configuración (Google colab. 2023)

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 79 de 128

- Acceso a *GPUs* sin coste adicional (Google colabouratory. 2023)
- Permite compartir contenido fácilmente (Google Colabouratory. 2023)

De acuerdo con García-García, P. (2022) Google Collaboratory es un entorno gratuito de Júpiter Notebook que no requiere configuración. En él se puede programar y ejecutar código Python en la nube, dando acceso a *GPUs* y permitiendo compartir contenido fácilmente. (García-García, P. 2022) Forma parte de la suite de aplicaciones de Google, almacenando todo lo desarrollado en el en el espacio de Google Drive.

Algunas de las principales características del *Colab* son:

- Dado que se ejecuta en una máquina de Google, no es necesario realizar ninguna configuración (García-García, P. 2022)
- Google proporciona acceso gratuito a las *GPU* (García-García, P. 2022)
- Es fácil de compartir, como cualquier archivo en el drive (García-García, P. 2022)

Además, ofrece una serie de cuadernos, donde se puede crear y ejecutar el código, siendo el entorno de trabajo en *Colab* (García-García, P. 2022). También permite combinar código ejecutable y texto enriquecido en un mismo documento, con imágenes, *HTML* o *LáTex*. Los cuadernos están compuestos por celdas independientes, las cuales son las unidades de ejecución que se proporcionan y donde el código va escrito, mostrando el resultado justo debajo de la misma. García-García, P. (2022) A pesar de ser independientes, todas las celdas se ejecutan utilizando el mismo *kernel*, guardando las variables declaradas de una celda a otra.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 80 de 128

Con *Google colab* se puede aprovechar toda la potencia de las bibliotecas de Python para analizar y visualizar datos. Además, se pueden importar datos desde una cuenta de Seguimiento (*tracking*) Google Drive o desde GitHub. También se puede cargar un conjunto de imágenes para entrenar un clasificador con dicho conjunto y evaluar modelos utilizando pocas líneas de código. Al ser ejecutado en la nube, se utiliza la potencia del hardware de Google, con sus *GPU* y *TPU*, independientemente de la potencia local que tenga el equipo (García-García, P. 2022) Por estos motivos es una de las herramientas más utilizadas en la comunidad del aprendizaje automático.

Sin embargo, tiene ciertas limitaciones, ya que la máquina virtual a la que se conecta tiene un tiempo máximo de conexión de 12 horas, además de que se puede desconectar un cuaderno si no se utiliza durante 90 minutos, lo que puede derivar en la pérdida de la ejecución realizada o de los datos generados en su utilización (García-García, P. 2022).

2.8.1. Restricciones.

De acuerdo con la información brindada en la página oficial de *Google colab* (2022) los recursos en *Colab* tienen prioridad para casos de uso interactivos. No se permiten las siguientes acciones en los entornos de ejecución de *Colab*:

- Hosting de archivos, entrega de contenido multimedia o cualquier oferta de servicios web no relacionada con el procesamiento interactivo con *Colab*.
- Descarga de *torrents* o intercambio de archivos entre pares.
- Utilización de un escritorio remoto o SSH.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 81 de 128

- Conexión a *proxies* remotos.
- Minería de criptomonedas.
- Ejecución de ataques de denegación del servicio.
- Robo de contraseñas.
- Utilización de varias cuentas para evitar las restricciones de acceso o el uso de recursos.
- Creación de *deepfakes*

2.9. Dataset

Un *Dataset* no es más que un conjunto de datos tabulados en cualquier sistema de almacenamiento de datos estructurado. El término Dataset hace referencia a una única base de datos de origen, la cual se puede relacionar con otras, cada columna del *Dataset* representa una variable y cada fila corresponde a cualquier dato que estemos tratando. (“¿Qué son los,” 2023)

Hay 4 géneros de *Datasets* clasificados según su origen y formato, que se usan según las pretensiones de los modelos de datos a trabajar: (Data Science. 2023)

- Fichero/Archivo: Un tipo de *dataset* que se caracteriza por ser seguro, rápido y eficiente, al poner a disposición un conjunto de datos en un archivador que es automático. Al momento de realizar el procesamiento, se accede al archivo correspondiente. (Talently. 2022)
- Carpeta: Aquí se almacenan diversos *datasets*, es decir, una gran cantidad de información, que, a su vez, se encuentran interconectados unos con otros. (Talently. 2022)

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 82 de 128

- Bases de datos: Estos *datasets* son utilizados en programas y aplicaciones específicas, dependiendo del tipo de formato que posean dichos datos. (Talently. 2022)
- Web: Son los datos que se almacenan en una página web y que normalmente se representan en URL, que contienen toda la información del sitio. (Talently. 2022).

2.10. Herramienta Make sense.

Make sense es una herramienta de etiquetado y anotación de datos que simplifica la creación y gestión de conjuntos de datos para el entrenamiento de modelos de aprendizaje automático.

El motor que impulsa las funcionalidades de IA es TensorFlow.js - versión JS del marco más popular para entrenar redes neuronales. Esta elección nos permite no solo acelerar su trabajo, sino también preocuparnos por la privacidad de sus datos, porque a diferencia de otras herramientas comerciales y de código abierto, sus fotos no tienen que ser transferidas al servidor (Skalski, P. 2023)

Esta herramienta presenta una interfaz intuitiva, que permite etiquetar imágenes, asignar categorías, asignando etiquetas precisas a cada imagen para identificar la presencia o ausencia de la Varroosis, y además de esto colaborar en proyectos de manera eficiente, ofreciendo opciones de exportación para integrarse con otras herramientas de aprendizaje automático (“How to Annotate Your Image,” 2020). Esta herramienta permite exportar las anotaciones en diferentes formatos como JSON, CSV, Yolo.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 83 de 128

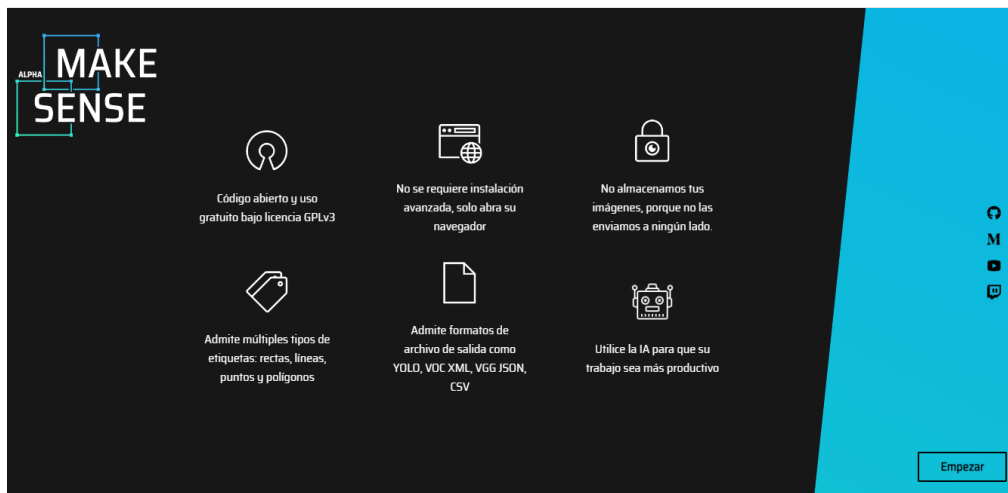


Imagen 22. Editor Make sense.

Make Sense (2023).

Capítulo 3. Metodología

En este capítulo se presenta la elección metodológica para desarrollar el sistema de detección de Varroosis en abejas mediante inteligencia artificial. Se basa en el enfoque del método CRISP-DM, y se explica cómo se aplicó este método junto con el modelo pre-entrenado YoloV5. Se detalla la transferencia de aprendizaje en el modelo para adaptarlo a la detección de Varroosis. También se describe la construcción del conjunto de datos, incluyendo la recopilación de imágenes etiquetadas de abejas afectadas y no afectadas.

3.1. Selección de metodología

Se realizó una aproximación metodológica con base a CRISP-DM ajustando cada una de las fases a los requerimientos propuestos excepto la fase de modelado la cual se reemplaza por la fase de entrenamiento. Esto debido a que en la fase de modelado se crea el modelo y en este caso se trabaja con un modelo pre-entrenado. Se trabajó en base a esta metodología debido a que ha

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 84 de 128

sido ampliamente reconocida como una de las mejores técnicas para proyectos de minería de datos y aprendizaje automático. En particular, su enfoque en resultados y su proceso altamente iterativo lo hacen ideal para la detección temprana de enfermedades en abejas mediante IA.

Este enfoque permite identificar rápidamente las mejores prácticas y estrategias para el uso efectivo de soluciones de aprendizaje automático en la detección temprana de enfermedades en abejas. Además, el proceso iterativo asegura que el modelo sea continuamente mejorado y perfeccionado para mantener su eficacia y organización en el tiempo (“IBM Documentation,” 2021). Además, CRISP-DM se enfoca en la generación de resultados concretos y reproducibles (“IBM Documentation,” 2021) lo que lo hace especialmente útil en proyectos de detección de enfermedades en abejas, donde es necesario contar con un alto grado de precisión y confiabilidad en los resultados obtenidos.

En conclusión, CRISP-DM es la mejor opción porque proporciona una estructura clara y está enfocada en resultados para la implementación de soluciones de IA y aprendizaje automático, así como un proceso altamente iterativo para la mejora continua del modelo de IA.

A continuación, se implementan las fases en el sistema de detección temprana de Varroosis en las abejas el cual resulta importante para garantizar que el proyecto se desarrolle de manera efectiva, mitigando los riesgos y mejorando la eficiencia, al obtener una idea más específica del modelo de detección de objetos a desarrollar.

3.1.1. Fases aplicadas al sistema de detección temprana de Varroosis.

1. Comprensión del problema:

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 85 de 128

La detección temprana de la Varroosis gracias a la aplicación de inteligencia artificial. La Varroosis es una enfermedad que ha trascendido al pasar de los años propagándose hasta los sitios más lejanos afectando a miles de apicultores en los diferentes sectores, convirtiéndose así en una gran amenaza. Por este motivo nace la necesidad de lograr detectar esta enfermedad temprana y asertivamente para evitar así mayores efectos negativos.

La aplicación de la inteligencia permite una detección más eficiente debido a puede procesar grandes cantidades de información proporcionadas en un tiempo reducido logrando crear soluciones que no son fácil de percibir por los humanos, de aquí nace la necesidad de realizar la recolección de información como imágenes, y contar con acceso a redes pre entrenadas que permitan integrar esta nueva información proporcionada creando así un sistema con nuevos objetivos.

2. Comprensión de los datos.

La recolección de imágenes reales de abejas se realizó en los apiarios que pertenecen a las zonas más cercanas, se aplicó una previa selección de la calidad de cada imagen identificando si cumple con los requerimientos establecidos, también se realiza la aplicación de una serie de procesos como lo es el etiquetado de cada una logrando captar las características más relevantes las cuales serán el integrante principal del entrenamiento de la red.

3. Preparación de los datos.

Luego de revisar cada una de las imágenes obtenidas se eliminan aquellas que no contaron con buena calidad. De esta manera se busca tener un conjunto de datos más acertado, organizándolo por medio de carpetas que permitan un mayor control entre imágenes y etiquetas.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 86 de 128

A continuación, se busca que las imágenes cumplan con los formatos establecidos, especialmente el tamaño aceptado por la red pre-entrenada escogida.

4. Modelado/Entrenamiento.

En esta fase se realizó el entrenamiento del modelo YoloV5x usando el conjunto de datos que en la fase anterior fueron debidamente preparados. El objetivo es mejorar el rendimiento del modelo ajustando los hiper parámetros. Al final de esta fase, el modelo de detección de objetos basado en YoloV5x estará listo para la detección temprana de la Varroosis de las abejas mediante la visión artificial.

5. Evaluación.

Una vez que se han desarrollado los modelos, es importante evaluar su rendimiento. En este caso, se evalúa la precisión del modelo YoloV5x en la detección de Varroa en abejas y esto se realiza midiendo la proporción de detecciones verdaderas en relación con todas las detecciones realizadas.

6. Implementación

El siguiente paso es implementar el modelo aplicado anteriormente en un entorno de producción. Esto puede incluir la integración de los modelos en una aplicación local o plataforma. También se realizan pruebas de aceptación del usuario para garantizar que el modelo sea fácil de usar y brinde los resultados esperados.

3.2. Selección del modelo pre entrenado YoloV5x

Se selecciona el modelo YoloV5x Large ya que es una opción adecuada para la detección temprana de Varroosis de abejas, esto porque es muy preciso y puede detectar objetos pequeños

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 87 de 128

en las imágenes. La Varroa es una enfermedad que infecta a las abejas y se manifiesta como ácaros que se adhieren a sus cuerpos. Estas garrapatas son muy pequeñas y difíciles de ver a simple vista, por lo que se necesita un medio de detección preciso y eficaz.

El YoloV5X-Large utiliza una arquitectura de red neuronal profunda entrenada previamente en una gran cantidad de imágenes para detectar objetos pequeños con alta precisión. Además, este modelo está optimizado para trabajar en tiempo real, por lo que es ideal para su uso en sistemas de seguimiento y detección precoz de Varroa en abejas.

Este modelo en especial nos ofrece características muy útiles al momento de realizar la detección como

- Velocidad: Este algoritmo mejora la velocidad de detección porque puede predecir objetos en tiempo real (Sojasingarayar, A. 2023).
- Alta precisión: Yolo es una técnica predictiva que proporciona resultados precisos con errores de fondo mínimos (Sojasingarayar, A. 2023).
- Fácil de instalar: YoloV5 solo requiere la instalación de antorcha y algunas bibliotecas ligeras de Python (Solawetz, J. 2023).
- Entrenamiento rápido: Los modelos YoloV5 entrenan extremadamente rápido, lo que ayuda a reducir los costos de experimentación a medida que construye su modelo (Solawetz, J. 2023).
- YoloV5 es pequeño. Específicamente, un archivo de pesos para YoloV5 es de 27 megabytes (Nelson, J. 2023).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 88 de 128

- Estructura intuitiva del sistema de archivos de datos: el diseño de la carpeta de archivos es intuitivo y fácil de navegar durante el desarrollo (Solawetz, J. 2023).

3.3 Transferencia de aprendizaje del modelo YoloV5x.

Para el desarrollo del sistema es importante tener en cuenta que este no se realiza desde cero, se aplica la transferencia de aprendizaje, debido a que luego de escoger el modelo *YoloV5* el cual viene pre-entrenado con un conjunto de datos ya establecido en este caso *Coco128* que contiene información sobre las imágenes, como su ruta de archivo, dimensiones.

Coco 128 además, también almacena las anotaciones de objetos presentes en las imágenes, incluyendo las categorías definidas para los objetos, este se toma como punto de partida y se ajusta (*fine-tuning*) en un nuevo conjunto de datos proporcionado para realizar la detección. Esto implica entrenar el modelo utilizando nuevas imágenes y anotaciones del conjunto de datos para realizar la detección.

De esta manera se mejora y se acelera el proceso de entrenamiento de *YoloV5*, ya que el modelo inicialmente tiene conocimientos previos y características aprendidas de los datos de entrenamiento masivo. Esto ayuda a reducir la cantidad de datos necesarios para entrenar el modelo desde cero y mejora la capacidad de generalización del modelo en el nuevo conjunto de datos, obteniendo así el prototipo del sistema de detección de Varroa en abejas. Aunque las pruebas se llevarán a cabo utilizando imágenes de abejas reales en lugar de un entorno real basado en

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 89 de 128

colmenas y sensores, el prototipo será capaz de detectar la presencia de Varroosis, permitiendo así convertirse en un base muy importante para fututos desarrollos.

3.3. Modelo de caso de uso del negocio

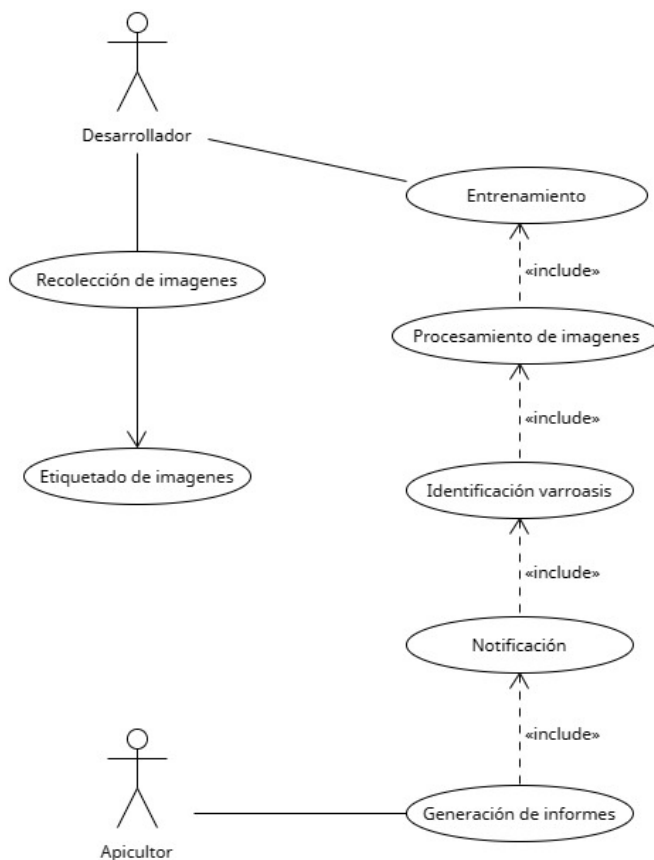


Imagen 23. Modelo de caso de uso.

Capítulo 4. Desarrollo e implementación

El presente capítulo se centra en el desarrollo e implementación del sistema de detección de Varroa en abejas utilizando el algoritmo YoloV5, conocido por su rapidez y precisión en la detección de objetos.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 90 de 128

A lo largo de este capítulo, se describirán los pasos clave llevados a cabo para construir el modelo, preparar el conjunto de datos, realizar el etiquetado de imágenes y definir la arquitectura del modelo. Además, se proporcionará una descripción detallada del algoritmo utilizado en este proceso.

Se describe la construcción del conjunto de datos, el etiquetado de imágenes y la arquitectura del modelo utilizado. También se proporciona una descripción general del algoritmo y su aplicación en la detección de Varroa en abejas.

Este capítulo sienta las bases para comprender el desarrollo y la implementación del sistema de detección de Varroa, proporcionando una visión general de los pasos realizados y los beneficios obtenidos al seleccionar YoloV5 como modelo de detección de objetos.

4.1. Construcción del conjunto de datos Dataset

En la construcción de la data set se tienen en cuenta diferentes procesos los cuales permiten realizar una detección óptima. Entre ellos se tiene la recolección de imágenes de abejas sanas e infestadas con Varroa, el total de imágenes corresponde a una cantidad de 3.000 de las cuales 2000 imágenes se tomaron específicamente en diferentes apiarios del sector y las 1000 imágenes restantes se obtuvieron de material publicado por Verastuchelova (2023) el cual contiene imágenes de abejas infestadas con Varroa en diferentes formas. Debido a que encontrar gran cantidad de imágenes de abejas infectadas con Varroa se torna un poco difícil se opta por tomar las imágenes publicadas por esta página.

Las imágenes se toman en diferentes colmenas, de esta forma se busca obtener material variado para ofrecer al sistema un mejor entrenamiento. Principalmente se busca que las imágenes

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 91 de 128

sean tomadas con una buena resolución, estas son tomadas con la cámara de celulares para optimizar el rendimiento en cuanto cantidad de imágenes que se toman. A continuación, se muestran algunos de los productos obtenidos:



Imagen 24. Foto de Abejas en una colmena.

Después de recolectar la cantidad de imágenes necesaria, se realiza una revisión intensa. Este paso es muy importante porque garantiza la calidad y consistencia de la imagen, lo que da como resultado datos más precisos durante el entrenamiento del modelo. Durante la limpieza, se realizan tareas para garantizar la calidad de la imagen. Se han eliminado imágenes borrosas o de baja calidad y se han corregido fallas como desenfoque, ruido o mala iluminación. Además, las imágenes se formatearon según fue necesario para garantizar un tamaño, una forma y una resolución uniformes en todo el conjunto de datos. El siguiente paso es verificar que cada imagen se está etiquetada de forma correcta y así cumplir con los objetivos establecidos anteriormente.

4.1.1 Características del Conjunto de Entrenamiento

- **Etiquetado:** cada imagen de la Dataset tenía que estar etiquetada correctamente mostrando la presencia y ubicación de las abejas, así como una clasificación de abejas sanas o

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 92 de 128

enfermas. Estas etiquetas son necesarias para enseñar al modelo a reconocer y diferenciar entre diferentes las diferentes opciones que se le presentan.

- Resolución y tamaño: las imágenes deben tener suficiente resolución y tamaño para permitir una identificación correcta. Casos diferentes: el conjunto de entrenamiento debe incluir una amplia gama de imágenes que representen diferentes escenarios y variaciones de la apariencia de abejas sanas y enfermas. Esto ayuda al modelo a generalizar e identificar patrones importantes en todas las condiciones.
- Equilibrio de clase: es importante que el conjunto de imágenes del Dataset contenga un número equilibrado de imágenes de abejas sanas y enfermas para evitar fallas en la detección. Esto asegura que el modelo aprenda correctamente las propiedades de estas dos clases y pueda hacer predicciones precisas en situaciones reales.
- Calidad y consistencia: las imágenes en el conjunto de entrenamiento deben ser de alta calidad y libres de ruidos o distorsiones que puedan dificultar la detección. Además, es fundamental que las funciones y características de las abejas sanas y enfermas permanezcan constantes en todas las imágenes para que el modelo aprenda de manera efectiva.

4.2. Etiquetado.

Para llevar a cabo el proceso de etiquetado, se preparó una carpeta dedicada que contenía todas las imágenes seleccionadas, y se ubicó en un lugar de fácil acceso al momento de cargarla en la herramienta *MakeSense*. Una vez que la carpeta, que albergaba el conjunto de datos, estuvo lista, se procedió a cargarla en la plataforma. El primer paso consistió en seleccionar los nombres

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 93 de 128

de las etiquetas, las cuales desempeñarán un papel crucial en la diferenciación de los distintos grupos de imágenes. En total, se emplearon tres etiquetas: "AbejaConVarroa", "AbejaSana" y "Varroa".

Posteriormente, se preparó la lista de imágenes, asegurándose de que estuvieran listas y disponibles para el proceso de etiquetado y anotación.

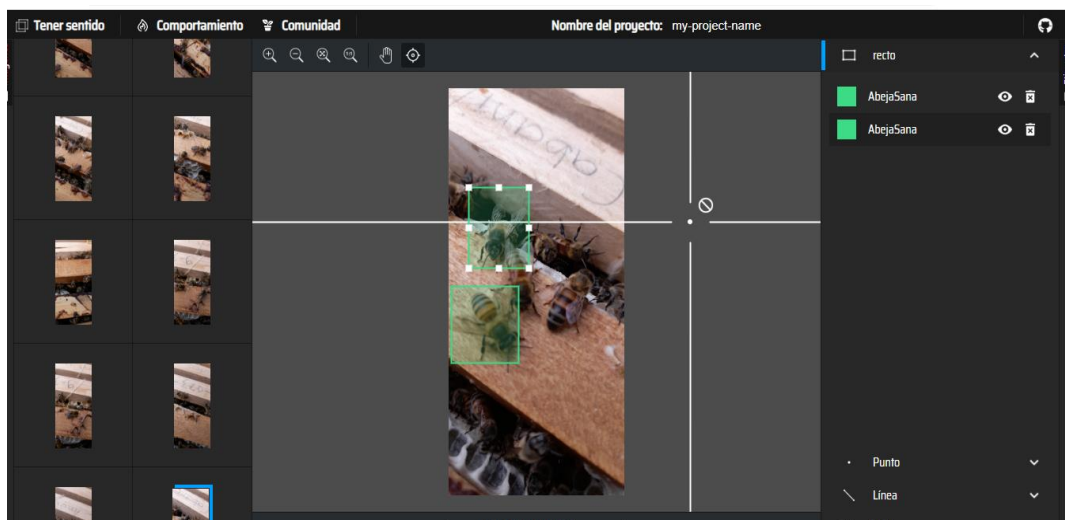


Imagen 25. Etiquetado en MakeSense.

El proceso de etiquetado implica seleccionar cada una de las imágenes como se puede observar en la imagen anterior en la herramienta *MakeSense*, para delimitar con un recuadro el elemento requerido, en este caso, las abejas presentes en la imagen. Luego, se asigna una de las etiquetas propuestas a cada recuadro. Es importante destacar que una imagen puede tener múltiples etiquetas, según los elementos presentes en ella. Este proceso se repite para cada una de las imágenes cargadas, asegurando así que todas sean debidamente etiquetadas y anotadas.

Una vez que todas las imágenes han sido debidamente etiquetadas, *MakeSense* ofrece la funcionalidad de exportar las etiquetas en diferentes formatos, permitiendo una integración

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 94 de 128

sencilla con diversos algoritmos y aplicaciones. En el caso particular de este proyecto, se ha seleccionado el formato Yolo, ya que será el algoritmo utilizado durante el proceso.

El formato Yolo es ampliamente utilizado en tareas de detección de objetos y proporciona una estructura específica para almacenar las coordenadas y las clases de los objetos detectados en cada imagen (A. Enrique, 2018). Al exportar las etiquetas en formato *Yolo*, se asegura una compatibilidad directa con el algoritmo y se facilita la implementación posterior del modelo de detección de Varroosis en las abejas (A. Enrique, 2018)

4.3. Arquitectura del Modelo

En términos de la estructura del modelo, YoloV5x se compone de varias capas convolucionales para extraer características relevantes de las imágenes y realizar las detecciones. Estas capas están organizadas de manera eficiente para lograr un equilibrio entre precisión y velocidad de detección (Wang et al., 2022). Además, se han establecido los parámetros del modelo, como la tasa de aprendizaje, el tamaño del lote (*batch-size*) y el número de épocas de entrenamiento, mediante experimentación y ajuste fino para optimizar el rendimiento del modelo en la detección de Varroosis en las abejas (Wang et al., 2022). Gracias a esta implementación, se espera obtener resultados más eficientes y precisos en la detección de Varroosis, lo que contribuirá a la salud y bienestar de las colonias de abejas.

A continuación, se observa la estructura de *YoloV5x*.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 95 de 128

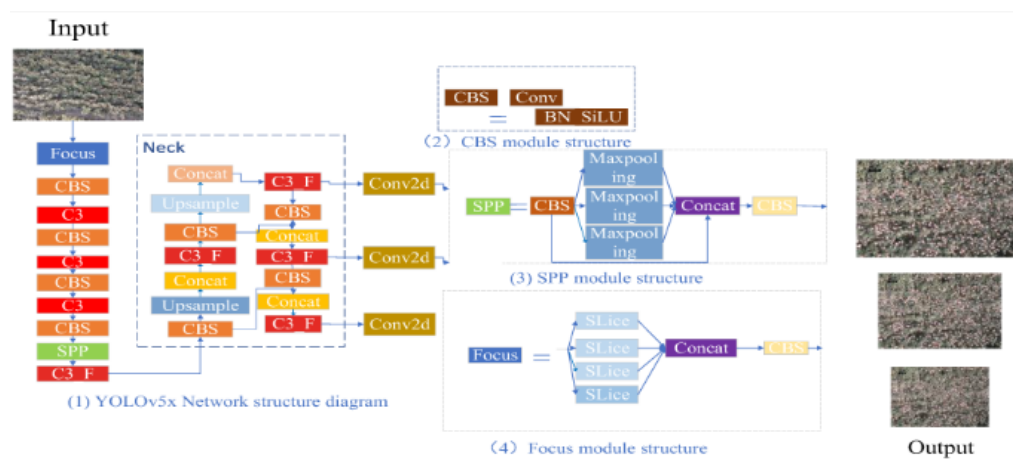


Imagen 26. Diagrama de estructura de la red YoloV5x.

Wang, X., Wang, X., Hu, C., Dai, F., Xing, J., Wang, E., ...Guo, W. (2022).

El modelo YoloV5x ha sido diseñado para mejorar la precisión y velocidad de detección en comparación con las versiones anteriores de *Yolo*. Utiliza un *backbone* o estructura principal que consta de diferentes módulos, como Focus, CBS y SPP, para extraer características relevantes de las imágenes y realizar la detección de objetos (Wang et al., 2022).

El módulo Focus divide el mapa de características de entrada y lo apila en dimensiones de canal, lo que reduce su tamaño y aumenta la cantidad de canales. Esto permite capturar información a diferentes escalas y niveles de detalle (Wang et al., 2022). El módulo CBS es el principal responsable de la extracción de características y se encarga de realizar convoluciones estándar, muestreo descendente y otras operaciones para identificar patrones relevantes en las imágenes. El módulo SPP (Spatial Pyramid Pooling) convierte el mapa de características de entrada en un tamaño fijo mediante una operación de convolución, lo que reduce la carga computacional del algoritmo (Wang et al., 2022).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 96 de 128

4.4. Entrenamiento de datos personalizados usando YoloV5x

Para el proceso de entrenamiento, se utilizó la distribución de YoloV5x. Para el proceso de entrenamiento se utilizó un ordenador con 16 GB de memoria, una tarjeta gráfica *NVIDIA GeForce GTX 1080 TI* y un procesador *Intel Core i7*. El proceso de entrenamiento fue relativamente sencillo, ya que, está muy bien explicado en la página de los autores.

Finalmente, después de 300 *epochs* (períodos de entrenamiento) se compararon los resultados entre ellos y se seleccionaron los coeficientes para la red neuronal generados a los 300 *epochs*. A partir del *dataset* creado con 3000 imágenes se inicia el entrenamiento.

Se llevaron a cabo los siguientes pasos

4.4.1. Creación de dataset yaml.

A continuación, se presenta el archivo de configuración del conjunto de datos, el cual establece la ubicación del directorio principal del conjunto de datos, así como las rutas relativas a los directorios de imágenes de entrenamiento (*train*), validación (*val*) y prueba (*test*). Además, también se incluye un diccionario de nombres de clases que se utiliza en el conjunto de datos.

Se crean las etiquetas,

1. Varroa,
2. abejaSana
3. abejaConVarroa

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 97 de 128

```

train: /content/data/images/train # train images (relative to 'path')
val: /content/data/images/val # val images (relative to 'path')
test: # test images (optional)

# Classes
names:
  0: varroa
  1: abejaSana
  2: abejaConVarroa

# Download script/URL (optional)
download: https://ultralytics.com/assets/data.zip

```

Imagen 27. Creación de etiquetas para entrenamiento.

Se determina usar la herramienta *Colab* (Google Colaboratory) para llevar a cabo el entrenamiento. Antes de comenzar, se aseguró de que la configuración de *GPU* estuviera activada en el cuaderno creado, lo que permite aprovechar el poder de cálculo de la *GPU* para acelerar el proceso de entrenamiento.

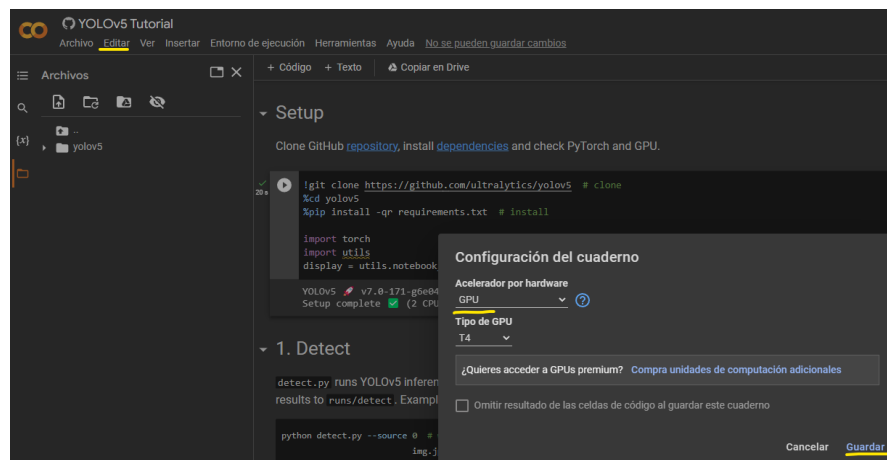


Imagen 28. Configuración del cuaderno de entrenamiento.

Ahora se debe clonar repositorio y requisitos de instalación.txt en un entorno *Python* $\geq 3.7.0$, incluido *PyTorch* ≥ 1.7 . Los modelos y conjuntos de datos se descargan automáticamente desde la última versión de *YoloV5*.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 98 de 128

Para clonar el repositorio desde GitHub se usa el comando,

git clone <https://github.com/ultralytics/YoloV5>

Cambiamos de directorio a la carpeta YoloV5 con el comando

cd YoloV5

Se organizan localmente las imágenes etiquetadas y no etiquetadas

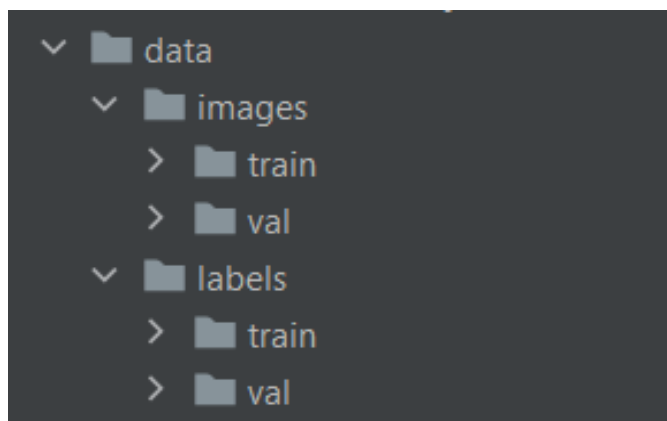


Imagen 29. Configuración del directorio de datos.

4.4.2. Inferencia

Se instalan los requisitos de instalación

pip install -r requirements.txt

Se importa la biblioteca *Torch*

Se importa el módulo *utils* y se configura el entorno del cuaderno para utilizar *YoloV5*.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 99 de 128

```

#clone YOLOv5 and
!git clone https://github.com/ultralytics/yolov5 # clone repo
%cd yolov5
%pip install -qr requirements.txt # install dependencies
%pip install -q roboflow

import torch
import os
from IPython.display import Image, clear_output # to display images

print(f"Setup complete. Using torch {torch.__version__} ({torch.cuda.get_device_properties(0).name if torch.cuda.is_available() else 'CPU'})")

Cloning into 'yolov5'...
remote: Enumerating objects: 15831, done.
remote: Total 15831 (delta 0), reused 0 (delta 0), pack-reused 15831
Receiving objects: 100% (15831/15831), 14.48 MiB | 25.83 MiB/s, done.
Resolving deltas: 100% (10873/10873), done.
/content/yolov5
184.3/184.3 kB 4.5 MB/s eta 0:00:00
592.7/592.7 kB 16.0 MB/s eta 0:00:00
62.7/62.7 kB 960.5 kB/s eta 0:00:00
56.3/56.3 kB 2.6 MB/s eta 0:00:00
58.8/58.8 kB 7.5 MB/s eta 0:00:00
67.8/67.8 kB 8.3 MB/s eta 0:00:00
Preparing metadata (setup.py) ... done
54.5/54.5 kB 6.9 MB/s eta 0:00:00
Building wheel for wget (setup.py) ... done
Setup complete. Using torch 2.0.1+cu118 (Tesla T4)

```

Imagen 30. Referencia comandos de clonación de repositorio YoloV5.

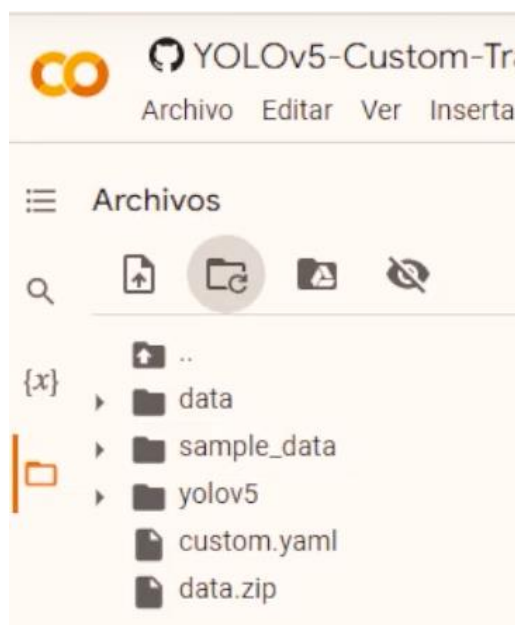


Imagen 31. Directorios en Google Colab de YoloV5.

Google Colaboratory. (2023, May 30)

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 100 de 128

En este punto, se carga el archivo comprimido con las imágenes para entrenamiento “data.zip”, tal como aparece referenciado en la imagen 3.9.

En seguida se descomprime con el comando

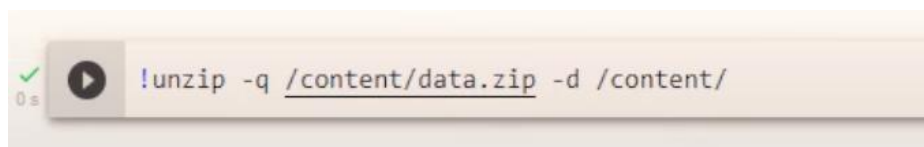


Imagen 32. Comando para descomprimir un archivo.

4.4.3. Entrenamiento.

Se organizan los directorios y se ejecuta el comando:

```
!python train.py --img 640 --batch 16 --epochs 300 --data /content/YoloV5/data/custom.yaml --weights YoloV5x.pt --cache
```

- *python train.py* indica que se ejecutará el script train.py el cual está diseñado para entrenar el modelo YoloV5.
- *img 640* especifica el tamaño de las imágenes, 640x640 píxeles.
- *epochs 300* indica el número de épocas o iteraciones completas que se realizan para entrenar el modelo.
- *data /content/YoloV5/custom.yaml*, en este archivo encontramos los datos de configuración de entrenamiento.
- *weights YoloV5x.pt*, se indica el tipo de modelo a usar.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 101 de 128

- Todos los resultados del entrenamiento se guardan con directorios de ejecución. Las pérdidas de entrenamiento y las métricas de rendimiento se guardan en *Tensorboard* y también en un archivo de registro.

4.4.4. Visualización.

Ejecutamos el siguiente comando,

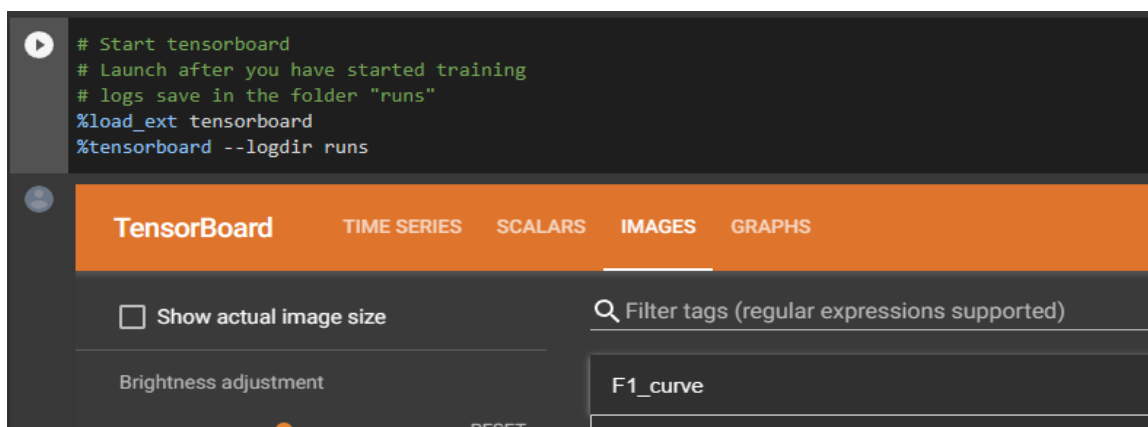


Imagen 33. Visualización del entrenamiento con TensorFlow.

Por último, se ejecuta el comando para exportar el modelo, que podrá ser usado en un entorno *PyCharm*.

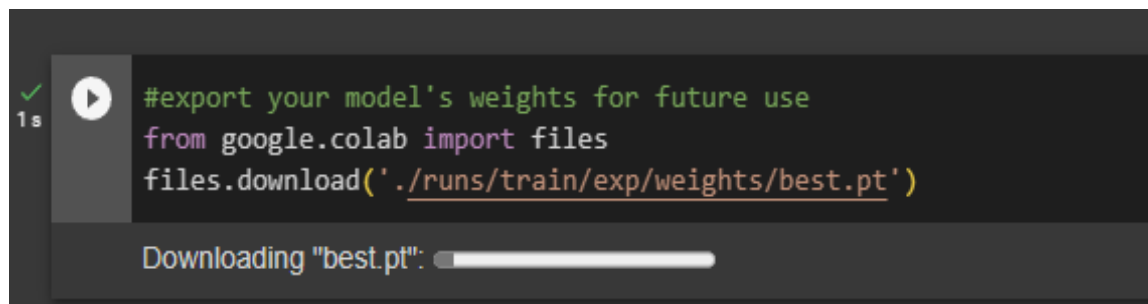


Imagen 34. Exportar modelos.

Google Colaboratory. (2023, May 30).

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 102 de 128

Ahora en colaboración con *PyCharm*, se ejecuta el siguiente comando en la consola

```
pip install -r
```

<https://raw.githubusercontent.com/ultralytics/YoloV5/master/requirements.txt>

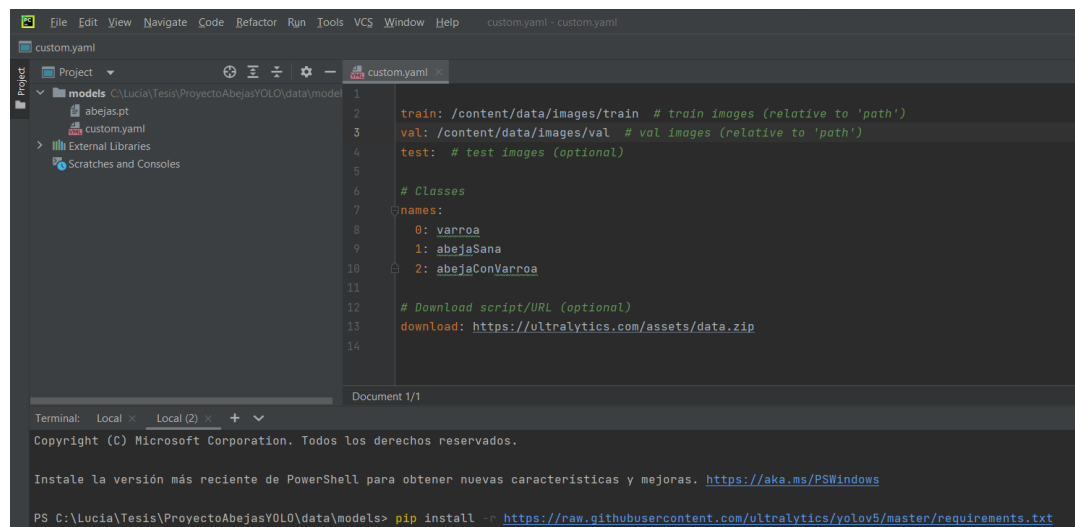


Imagen 35. Instalación de requerimientos en entorno local PyCharm.

Desde el entorno local, en *pyCharm* se carga un modelo *YoloV5* pre-entrenado que realiza detecciones en tiempo real desde una cámara y muestra los resultados en una ventana de visualización.

1. Importamos las bibliotecas necesarias: *torch*, *cv2* (*OpenCV*) y *numpy* (para tiempo real).
2. Se lee el modelo pre-entrenado: Se carga el modelo *YoloV5* personalizado desde la ubicación especificada en el parámetro *path*. Esto se realiza utilizando *torch.hub.load()* del repositorio '*ultralytics/YoloV5*'. El modelo se almacena en la variable *model*.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 103 de 128

3. Se realizó la captura de video: Se crea un objeto *VideoCapture* utilizando el método `cv2.VideoCapture(0)`, que abre la cámara principal del dispositivo.
4. Se coloca bucle *while* para que se ejecuten de manera indefinida hasta que nosotros queramos
5. Realización de las detecciones: Se pasa el *frame* al modelo *YoloV5* utilizando *model (frame)*. Esto devuelve un objeto de detección que contiene información sobre los objetos encontrados en el *frame*.
6. Obtención de información de detección: Se utiliza *detect.pandas().xyxy[0]* para obtener la información de detección en formato pandas *DataFrame*. Esta variable *info* contiene las coordenadas (x, y) de las esquinas de los objetos detectados, así como sus etiquetas y puntuaciones de confianza.
7. Mostrar el *frame* con las detecciones: Se muestra el *frame* actualizado con las detecciones utilizando *np.squeeze(detect.render())* y *cv2.imshow()*. Esto muestra una ventana con el nombre 'Detector de Abejas' que muestra el *frame* con las cajas delimitadoras y las etiquetas de los objetos detectados.
8. Se utiliza *cv2.waitKey(5)* para esperar 5 milisegundos y leer el teclado. Si se presiona la tecla 'Esc' (código 27), se rompe el bucle y se finaliza el programa.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 104 de 128

```

1  # Importamos librerías
2  import torch
3  import cv2
4  import numpy as np
5
6  # Leemos el modelo
7  model = torch.hub.load('ultralytics/yolov5', 'custom', trust_repo=True,
8                        path = 'C:/Lucia/Tesis/ProyectoAbejasYOL0/data/models/abejas.pt')
9
10 # Realizo Videocaptura
11 cap = cv2.VideoCapture(0)
12
13 # Empezamos
14 while True:
15     # Realizamos lectura de frames
16     ret, frame = cap.read()
17
18     # Correccion de color
19     #frame = cv2.cvtColor(frame, cv2.COLOR_BGR2RGB)
20
21     # Realizamos las detecciones
22     detect = model(frame)
23
24     info = detect.pandas().xyxy[0] # im1 predictions
25     print(info)
26
27     # Mostramos FPS
28     cv2.imshow('Detector de Abejas', np.squeeze(detect.render()))
29
30     # Leemos el teclado
31     t = cv2.waitKey(5)
32     if t == 27:
33         break
34
35 cap.release()
36 cv2.destroyAllWindows()

```

Imagen 36. Personalización del modelo en entorno local PyCharm.

Capítulo 5. Resultados

En este capítulo se presentan los resultados obtenidos del sistema de detección de Varroa en abejas utilizando el algoritmo YoloV5. Se analizan dos aspectos clave: la inferencia del modelo y las validaciones realizadas.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 105 de 128

Los resultados obtenidos en este capítulo son fundamentales para evaluar el rendimiento y la eficacia del sistema de detección de Varroa en abejas. Estos resultados respaldan la elección de *YoloV5* como modelo de detección y proporcionan una evaluación objetiva de su capacidad para detectar Varroa con precisión y confiabilidad.

En resumen, este capítulo presenta los resultados detallados de la inferencia y las validaciones realizadas en el sistema de detección de Varroa. Estos resultados son cruciales para comprender el desempeño del sistema y su utilidad en la detección temprana de la Varroosis en abejas.

Epoch	GPU_mem	box_loss	obj_loss	cls_loss	Instances	Size	
7/299	14.6G	0.05569	0.05041	0.003115	11	640: 100% 24/24	[00:23<00:00, 1.03it/s]
	Class	Images	Instances	P	R	mAP50	mAP50-95: 100% 1/1 [00:00<00:00, 1.13it/s]
	all	29	116	0.609	0.563	0.598	0.214
Epoch	GPU_mem	box_loss	obj_loss	cls_loss	Instances	Size	
8/299	14.6G	0.05234	0.0538	0.002665	27	640: 100% 24/24	[00:23<00:00, 1.03it/s]
	Class	Images	Instances	P	R	mAP50	mAP50-95: 100% 1/1 [00:00<00:00, 1.24it/s]
	all	29	116	0.616	0.543	0.606	0.217
Epoch	GPU_mem	box_loss	obj_loss	cls_loss	Instances	Size	
9/299	14.6G	0.05065	0.04832	0.002138	13	640: 100% 24/24	[00:23<00:00, 1.03it/s]
	Class	Images	Instances	P	R	mAP50	mAP50-95: 100% 1/1 [00:00<00:00, 1.24it/s]
	all	29	116	0.6	0.526	0.577	0.203
Epoch	GPU_mem	box_loss	obj_loss	cls_loss	Instances	Size	
10/299	14.6G	0.0485	0.04615	0.002408	6	640: 100% 24/24	[00:23<00:00, 1.03it/s]
	Class	Images	Instances	P	R	mAP50	mAP50-95: 100% 1/1 [00:00<00:00, 1.25it/s]
	all	29	116	0.652	0.534	0.596	0.238
Epoch	GPU_mem	box_loss	obj_loss	cls_loss	Instances	Size	
11/299	14.6G	0.04557	0.04581	0.001935	106	640: 75% 18/24	[00:18<00:06, 1.00s/it]

Imagen 37. Entrenamiento de epochs en Google Colab.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 106 de 128

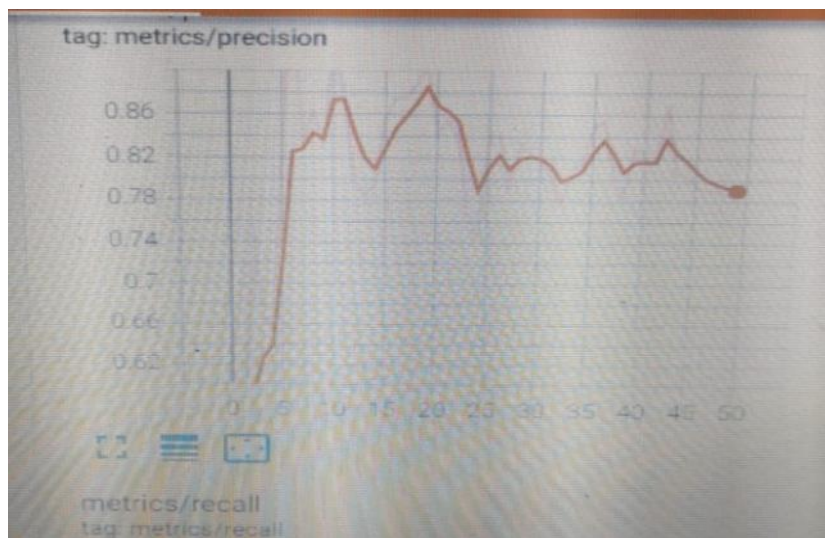


Imagen 38. Métrica de precisión epochs en Google Colab.

El gráfico tags:metrics/precisión muestra que el algoritmo de entrenamiento produce resultados con valores de precisión superiores a 0,80, llegando a sobrepasar el 0.86. Esto significa que el modelo entrenado logra una alta precisión en la detección de objetos. Un valor de precisión superior a 0,80 indica que el modelo pudo identificar correctamente más del 80% de los objetos en la imagen de evaluación. De esta manera, al tener alta precisión a la hora de detectar abejas o Varroas, brinda una alta confiabilidad para saber si una colmena está infectada o no por Varroa.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 107 de 128

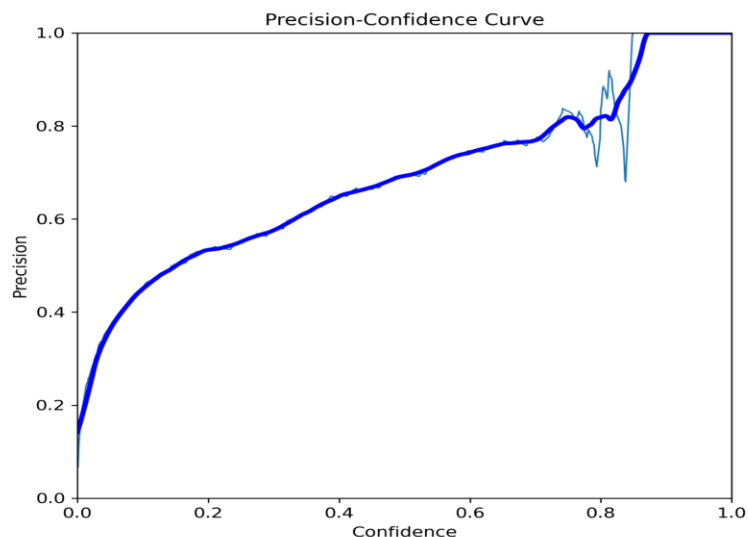


Imagen 39. Gráfica de Precision-Confidence Curve.

Esto muestra que, al evaluar los conjuntos de datos, todas las clases de objetos en la imagen, tal como Varroa o abeja, se encuentran correctamente con una precisión del 100%, cuando el modelo YoloV5 se detecta con una confianza de 0.80 o mayor. En otras palabras, el modelo puede identificar y clasificar todas las categorías diferentes de objetos en la imagen con gran precisión.



Imagen 40. Resultado detección abeja.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 108 de 128

Finalmente, la prueba de detección se realiza utilizando el algoritmo Diseñado Se puede evidenciar que detecta correctamente las abejas de las imágenes dadas.

Estos resultados muestran que el algoritmo de detección realizado puede identificar abejas en imágenes con alta precisión. Por lo tanto, los resultados de los gráficos anteriores confirman el rendimiento del modelo de detección presentado. Estos resultados son positivos y sugieren que los algoritmos de detección pueden usar efectivamente las aplicaciones relacionadas con la identificación y el monitoreo de las abejas en los medios de apicultura.

Para la detección especial de abejas, los resultados obtenidos con YoloV5 fueron eficientes y precisos. El método permitió una detección eficiente y precisa de las abejas, lo que proporcionó información valiosa para su investigación y conservación. Con los recursos disponibles, los usuarios pueden aprovechar al máximo las características de YoloV5 y obtener resultados de alta calidad en sus proyectos de identificación de abejas.

Capítulo 6. Conclusiones y Trabajos Futuros

6.1. Conclusiones

En conclusión, este estudio logró alcanzar los objetivos planteados de fortalecer los mecanismos de detección de *Varroosis destructor* en las abejas mediante el desarrollo de un algoritmo de inteligencia artificial. Los principales resultados se presentan a continuación:

- El algoritmo de inteligencia artificial ha demostrado ser una herramienta eficaz y prometedora para la detección temprana de *Varroosis destructor* en abejas. Su capacidad

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 109 de 128

para detectar el estado de las abejas a través de imágenes permitió la identificación precisa y eficiente de las abejas enfermas.

- La recopilación de datos fotográficos para crear el material fue esencial para entrenar y validar el algoritmo. La variedad de imágenes recopiladas aumentó la precisión y la generalización del modelo, lo que garantiza resultados confiables en diferentes escenarios y condiciones.
- El análisis exhaustivo de los datos de Varroosis proporcionó una comprensión más profunda de la enfermedad y su impacto en las colonias de abejas. Esto hizo posible identificar características específicas que son importantes para la detección precisa y oportuna y el diseño de algoritmos de IA.
- La identificación de la presencia de *Varroa destructor* en colonias de abejas en el macizo colombiano resaltó la importancia de contar con herramientas efectivas de detección. El algoritmo desarrollado ha demostrado su capacidad para identificar abejas enfermas.

6.2 Trabajos Futuros.

Un enfoque prometedor en el trabajo futuro es la implementación local de un sistema de detección de Varroa en abejas utilizando inteligencia artificial. Esto significa cumplir con todos los requisitos de hardware y software para entrenar y operar una red neuronal. La aplicación de este enfoque a nivel local lograría una mayor autonomía del proceso de seguimiento y descubrimiento que permite una adaptación precisa a los requisitos específicos del entorno apícola. Además, la implementación local garantiza la usabilidad del sistema y la capacidad de respuesta

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 110 de 128

en tiempo real, esenciales para la detección temprana y la intervención oportuna de infecciones de Varroa.

Adicionalmente se considera la posibilidad de completar el proyecto ampliando el material utilizado en el entrenamiento de la red neuronal Convolutiva. Aumentar el número de imágenes almacenadas en el Dataset sería muy útil para mejorar la precisión y la capacidad de detección de la Varroa de las abejas. Tal expansión del conjunto de datos significa la recopilación de diferentes imágenes que cubren diferentes entornos, condiciones ambientales y niveles de infestación de Varroa.

Además, se debe realizar un etiquetado cuidadoso y preciso de estas imágenes para garantizar un entrenamiento más completo del modelo. Al aumentar la variedad y el número de muestras en el material, aumentaría la probabilidad de obtener una identificación precisa y confiable en diferentes situaciones. Por otro lado, se destaca la importancia de promover la colaboración y el intercambio de datos entre investigadores y organizaciones, lo que enriquecería los datos disponibles y promovería avances significativos en este campo de investigación.

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 111 de 128

Referencias

A, enrique. (2018, May 12). Detección de objetos con Yolo: implementaciones y como usarlas.

Retrieved May 22, 2023, from Medium website:

<https://medium.com/@enriqueav/detecci%C3%B3n-de-objetos-con-Yolo-implementaciones-y-como-usarlas-c73ca2489246>

Abejas - características, diferencia con las avispas y por qué son importantes. (2023). Retrieved

June 24, 2023, from www.nationalgeographic.com.es website:

<https://www.nationalgeographic.com.es/animales/abeja>

Academia Win. (2023). Descubre las ventajas y desventajas del método Canvas en negocios.

Academia Win. Retrieved from <https://academiawin.com/ventajas-y-desventajas-de-canvas>

AcademiaLab. (2023, May 26). Procesamiento de imágenes _ Retrieved from [https://academia-](https://academia-lab.com/enciclopedia/procesamiento-de-imagenes)

[lab.com/enciclopedia/procesamiento-de-imagenes](https://academia-lab.com/enciclopedia/procesamiento-de-imagenes)

Adith Narein, T. (2021). Single Shot Detector (SSD) + Architecture of SSD. OpenGenus IQ:

Computing Expertise & Legacy. Retrieved from <https://iq.opengenus.org/single-shot-detector>

Algoritmo de detección de objetivos basado en aprendizaje profundo: SSD - programador clic.

(2023, May 24). Retrieved from <https://programmerclick.com/article/6461150291>

Apis mellifera: características, hábitat, reproducción, alimentación. (2020, December 18).

Retrieved from <https://www.lifeder.com/apis-mellifera>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 112 de 128

Árboles de decisión: qué son y cuál es su uso en Big Data. (2023, April 17). Retrieved from

<https://www.unir.net/ingenieria/revista/arboles-de-decision>

Asale, R.-, & Rae. (2023, June 02). algoritmo | Diccionario de la lengua española. Retrieved

from <https://dle.rae.es/algoritmo?m=form>

Asale, R.-, & Rae. (2023, June 02). interfaz | Diccionario de la lengua española. Retrieved from

<https://dle.rae.es/interfaz>

Asale, R.-, & Rae. (2023, June 02). polinización | Diccionario de la lengua española. Retrieved

from <https://dle.rae.es/polinizaci%C3%B3n>

Asana. (2023, May 21). Asana. Retrieved from [https://asana.com/es/resources/decision-tree-](https://asana.com/es/resources/decision-tree-analysis)

[analysis](https://asana.com/es/resources/decision-tree-analysis)

Barcelona Geeks. (2023 May 26)Conceptos básicos del procesamiento de imágenes

digitales. Retrieved from [https://barcelongeeks.com/conceptos-basicos-del-procesamiento-](https://barcelongeeks.com/conceptos-basicos-del-procesamiento-de-imagenes-digitales)

[de-imagenes-digitales](https://barcelongeeks.com/conceptos-basicos-del-procesamiento-de-imagenes-digitales)

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. (2020). YOLOv4: Optimal Speed and Accuracy of

Object Detection. Retrieved from <https://arxiv.org/pdf/2004.10934.pdf>

Bounous, C., & Boga, V. (2005). Fundamentos para el control de Varroa y loque americana.

Montevideo, Uruguay, 7.

<http://www.inia.uy/Publicaciones/Documentos%20compartidos/111219240807162010.p>

[df](http://www.inia.uy/Publicaciones/Documentos%20compartidos/111219240807162010.p)Cao, L., Zheng, X., & Fang, L. (2023). The Semantic Segmentation of Standing Tree

Images Based on the YOLO V7 Deep Learning Algorithm. 12(4), 929–929.

<https://doi.org/10.3390/electronics12040929>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 113 de 128

Cánovas Alcázar, J. (n.d.). VII Congreso SEAE Zaragoza 2006 No 6 VARROA (*Varroa jacobsoni*) SITUACIÓN ACTUAL Y MÉTODOS DE CONTROL. Retrieved from

<https://www.agroecologia.net/recursos/publicaciones/publicaciones-online/2006/CD%20Congreso%20Zaragoza/Ponencias/6%20C%C3%A1novas%20Com-%20Varroa.pdf>

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T.P., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*.

Claudia. (2022). *idUS - Depósito de Investigación Universidad de Sevilla*. *Idus.us.es*.
<https://hdl.handle.net/11441/140536>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 114 de 128

Colaboradores de los proyectos Wikimedia. (2023, May 12). Repositorio (contenido digital) -

Wikipedia, la enciclopedia libre. Retrieved from

[https://es.wikipedia.org/w/index.php?title=Repositorio \(contenido digital\)&oldid=151117529](https://es.wikipedia.org/w/index.php?title=Repositorio_(contenido_digital)&oldid=151117529)

CRISP-DM: una metodología para minería de datos en salud - healthdataminer.com. (2019,

November 04). Retrieved from <https://healthdataminer.com/data-mining/crisp-dm-una-metodologia-para-mineria-de-datos-en-salud>

Dan. (2023, March 23). ¿Qué es la inteligencia artificial (IA)? - IAfuturAI. Retrieved May 16,

2023, from IAfuturAI website: <https://iafuturai.com/que-es-la-inteligencia-artificial/>

Data Mining using CRISP-DM methodology. (2023, April 07). Retrieved from

<https://www.section.io/engineering-education/data-mining-using-crisp-dm-methodology>

De Jeu Boronat, C. (2022). Estudio y comparación de las versiones de Yolo para la aplicación

al tenis en tiempo real. (Trabajo Fin de Grado Inédito). Universidad de Sevilla, Sevilla.

Retrieved from

Diwan, T., G Anirudh, & Tembhurne, J. V. (2022). Object detection using Yolo: challenges,

architectural successors, datasets and applications. <https://doi.org/10.1007/s11042-022-13644-y>

Efren, R., & Jorge. (2016). Identificación y caracterización de abejas (Apis mellifera) resistente

a la Varroosis. Agrosavia.co, 105–117. <http://hdl.handle.net/20.500.12324/21100>

Equipo de redacción. (2022, October 14). Abeja. Retrieved June 24, 2023, from

bioenciclopedia.com website: <https://www.bioenciclopedia.com/abeja-156.html>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 115 de 128

Facundo, Q., Laura, L., De, U., Ronchetti, S., Lanzarini, F., & Cristina, L. (2019). PALABRAS

CLAVE: Machine Learning, Visión por Computador, Reconocimiento de gestos.

Investigación Joven, 6. Retrieved from

http://sedici.unlp.edu.ar/bitstream/handle/10915/95750/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y

Farib Khondoker, Rao, S. V., Spanias, A., & Cihan Tepedelenlioglu. (2018). Photovoltaic Array Simulation and Fault Prediction via Multilayer Perceptron Models.

<https://doi.org/10.1109/iisa.2018.8633699>

Farib Khondoker, Rao, S. V., Spanias, A., & Cihan Tepedelenlioglu. (2018). Photovoltaic Array Simulation and Fault Prediction via Multilayer Perceptron Models. International Conference on Information, Intelligence, Systems and Applications.

<https://doi.org/10.1109/iisa.2018.8633699>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 116 de 128

Faster R CNN ML. (2022, September 16). GeeksforGeeks. Retrieved from
<https://www.geeksforgeeks.org/faster-r-cnn-ml>

Figure 1.1: The Knowledge Discovery in Databases (KDD) process (Fayyad... (2019). Retrieved May 19, 2023, from ResearchGate website: https://www.researchgate.net/figure/The-Knowledge-Discovery-in-Databases-KDD-process-Fayyad-et-al-1996_fig1_220673065

Flores Velásquez, D. A. (2019). Técnicas de machine learning aplicado a la predicción de fuga de clientes en una institución de salud previsual bajo la metodología KDD (Doctoral dissertation, Universidad Andrés Bello).

Fois, G., Agüero Crovella, G. A., & Britos, P. V. (2020). Evaluación comparativa de las metodologías Team Data Science Process TDSP y Analytics Solutions Unified Method for Data Mining ASUM-DM desde la perspectiva de la ciencia de datos.

García-García, P. (2022). Sistema de seguimiento de personas desde un dron. Rua.ua.es.
<http://hdl.handle.net/10045/125320>

Gil, D. Á. (2021). Metodología CRISP-DM - Adictos al trabajo Tutoriales. Adictos al trabajo. Retrieved from <https://www.adictosaltrabajo.com/2021/01/14/metodologia-crisp-dm>

Gonzalez, L. (2020). Ventajas y Desventajas de los Algoritmos de Regresión - 🤖 Aprende IA. Retrieved from <https://aprendeia.com/ventajas-y-desventajas-de-los-algoritmos-de-regresion-machine-learning>

Google Colab. (2022, November 04). Retrieved from
<https://research.google.com/colaboratory/faq.html>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 117 de 128

Google Colaboratory. (2023, May 23). Retrieved from

https://colab.research.google.com/#scrollTo=5fCEDCU_qrC0

Grosso, G. S. (2014). Presencia e incidencia forética de Varroa destructor A.(Mesostigma:

Varroidae) en colonias de abejas Apis mellifera (Hymenoptera: Apidae), en Colombia.

Zootecnia Tropical, 30(2). [https://www.researchgate.net/profile/Guillermo-Salamanca-](https://www.researchgate.net/profile/Guillermo-Salamanca-Grosso/publication/235993479_ZTF-Varroa/links/0deec51550562a3d2c000000/ZTF-Varroa.pdf)

[Grosso/publication/235993479_ZTF-Varroa/links/0deec51550562a3d2c000000/ZTF-](https://www.researchgate.net/profile/Guillermo-Salamanca-Grosso/publication/235993479_ZTF-Varroa/links/0deec51550562a3d2c000000/ZTF-Varroa.pdf)

[Varroa.pdf](https://www.researchgate.net/profile/Guillermo-Salamanca-Grosso/publication/235993479_ZTF-Varroa/links/0deec51550562a3d2c000000/ZTF-Varroa.pdf) (Accessed: 16 May 2023).

Grosso, S., Patricia, & Arias, R. (2013). Presencia e incidencia forética de Varroa destructor A.

(Mesostigma: Varroidae) en colonias de abejas Apis mellifera (Hymenoptera: Apidae),

en Colombia. Zootecnia Tropical, 30(2), 183–195. Retrieved from

http://ve.scielo.org/scielo.php?pid=S0798-72692012000200007&script=sci_arttext

Grosso, S., Patricia, & Arias, R. (2013). Presencia e incidencia forética de Varroa destructor A.

(Mesostigma: Varroidae) en colonias de abejas Apis mellifera (Hymenoptera: Apidae),

en Colombia. Zootecnia Tropical, 30(2), 183–195. Retrieved from

http://ve.scielo.org/scielo.php?pid=S0798-72692012000200007&script=sci_arttext

Guerrero Paredes, M. D. P. (2017). Aplicación De La Metodología Canvas Para La Creación

De Una Pollería Pollo Dorado En La Ciudad De Pacora Chiclayo 2017.

[https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/33201/guerrero_pm.pdf?se-](https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/33201/guerrero_pm.pdf?sequence=1&isAllowed=y)

[quence=1&isAllowed=y](https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/33201/guerrero_pm.pdf?sequence=1&isAllowed=y)

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 118 de 128

Han, C. (2019, January 3). 5 Steps of a Data Science Project Lifecycle - Towards Data Science.

Retrieved May 19, 2023, from Medium website: <https://towardsdatascience.com/5-steps-of-a-data-science-project-lifecycle-26c50372b492>

Haya, P. (2022). La metodología CRISP-DM en ciencia de datos - IIC. Instituto de Ingeniería del Conocimiento. Retrieved from <https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos>

Haya, P. (2022). La metodología CRISP-DM en ciencia de datos - IIC. Instituto de Ingeniería del Conocimiento. Retrieved from <https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos>

Hermitaño Castro, J.A. (2022). Aplicación de Machine Learning en la Gestión de Riesgo de Crédito Financiero: Una revisión sistemática. Interfaces.

Honeybee.jpg. (2022, May 26). Retrieved June 24, 2023, from National Geographic website: <https://www.nationalgeographic.com/animales/abeja-melifera>.

Hotz, N. (2023). OSEMN Data Science Life Cycle - Data Science Process Alliance. Data Science Process Alliance. Retrieved from <https://www.datascience-pm.com/osemn>

How to Annotate Your Images Using the MakeSense Tool. (2020). Retrieved May 22, 2023, from Xailient.com website: <https://xailient.com/blog/how-to-annotate-your-images-using-the-makesense-tool/>

How to Create Infrastructure for Data Science Projects | .cult by HoneyPot. (2021). Retrieved May 2, 2023, from HoneyPot.io website: <https://cult.honeypot.io/reads/infrastructure-data-science-projects/>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 119 de 128

IBM Documentation. (2021, August 17). Retrieved from <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>

IBM Documentation. (2021, August 17). Retrieved from <https://www.ibm.com/docs/es/spss-modeler/saas?topic=modeling-overview>

IBM Documentation. (2021, August 17). Retrieved from <https://www.ibm.com/docs/es/spss-modeler/saas?topic=evaluation-overview>

IBM Documentation. (2021, August 17). Retrieved from <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>

Inteligencia artificial: la cuarta revolución industrial. (2017). Retrieved May 16, 2023, from

Google Books website:

https://books.google.com.mx/books?hl=es&lr=&id=sR3NDwAAQBAJ&oi=fnd&pg=PT2&dq=la+inteligencia+artificial+d%C3%A9bil+y+la+inteligencia+artificial+fuerte&ots=oj-TvhZGX0&sig=wIQ-3LtRtvp_p0ZI8kq65mQ04qg#v=onepage&q=la%20inteligencia%20artificial%20d%C3%A9bil%20y%20la%20inteligencia%20artificial%20fuerte&f=false

Jay. (2022, May 6). Introducción a Yolo: detección de objetos en tiempo real - HashDork.

Retrieved May 20, 2023, from HashDork website: <https://hashdork.com/es/Yolo/>

Knowledge Discovery in Databases. (2017, August 18). Retrieved from

<https://www.techopedia.com/definition/25827/knowledge-discovery-in-databases-kdd>

Larispardo. (2022). ¿Qué es y para qué sirve una regresión lineal en machine learning? Platzi.

Retrieved from <https://platzi.com/blog/que-es-regresion-lineal>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 120 de 128

Latam. (2020). Google Colab: ¿qué es y cómo usarlo? Alura. Retrieved from

<https://www.aluracursos.com/blog/google-colab-que-es-y-como-usarlo>

Lavo Córde Nas Mattíneçl, G. (n.d.). v LA Varroosis Situación en Colornbia. Retrieved from

https://repository.agrosavia.co/bitstream/handle/20.500.12324/21091/39072_22390.pdf?sequence=1&isAllowed=y

Lerch, D. (2021). Deep Learning con redes pre-entrenadas en ImageNet - Neuron4 - Medium.

Medium. Retrieved from <https://medium.com/neuron4/redes-pre-entrenadas-en-imagenet-30d858c37b1f>

likebupt. (2023, May 21). Regresión lineal: referencia de componente - Azure Machine

Learning. Retrieved from <https://learn.microsoft.com/es-es/azure/machine-learning/component-reference/linear-regression>

López Pacheco, M. A. (2017). Identificación de sistemas no lineales con redes neuronales

convolucionales (Master's thesis, Tesis (MC)--Centro de Investigación y de Estudios Avanzados del IPN Departamento de Control Automático). Recuperado 20 de mayo de 2023, de Cinvestav.mx website:

<https://repositorio.cinvestav.mx/bitstream/handle/cinvestav/2207/SSIT0014857.pdf>

luisquintanilla. (2023, June 02). Tutorial: Detección de objetos mediante un modelo de

aprendizaje profundo de ONNX - ML.NET. Retrieved from

<https://learn.microsoft.com/es-es/dotnet/machine-learning/tutorials/object-detection-onnx>

Machine learning: Trends, perspectives, and prospects. (2015). Retrieved May 3, 2023, from

Science website: <https://www.science.org/doi/10.1126/science.aaa8415>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 121 de 128

Manual APA 6a edición | Pontificia Universidad Javeriana, Cali. (2023, June 02). Retrieved from <https://www.javerianacali.edu.co/centro-escritura/recursos/manual-apa-6a-edicion>

*María, R., & Ferrari, S. (2016). ANÁLISIS TECNOLÓGICOS Y PROSPECTIVOS
SECTORIALES PROSPECTIVA TECNOLÓGICA AL 2025 DEL COMPLEJO APÍCOLA.
Retrieved from
https://www.argentina.gob.ar/sites/default/files/2_agro_atypsii_apicola.pdf*

*MarioMassaro. (2023). GitHub - MarioMassaro/Object-Deteccion-With-Yolo: Cree una guía
para el uso de YoloV5 para modelos personalizados de detección de objetos a tiempo
real. Retrieved May 21, 2023, from GitHub website:
<https://github.com/MarioMassaro/Object-Deteccion-With-Yolo>*

*marktab. (2023). ¿Qué es el Proceso de ciencia de datos en equipo (TDSP)? - Azure
Architecture Center. Retrieved May 19, 2023, from Microsoft.com website:
<https://learn.microsoft.com/es-es/azure/architecture/data-science-process/overview>*

*marktab. (2023, May 17). ¿Qué es el Proceso de ciencia de datos en equipo (TDSP)? - Azure
Architecture Center. Retrieved from [https://learn.microsoft.com/es-es/](https://learn.microsoft.com/es-es/azure/architecture/data-science-process/overview)
[es/azure/architecture/data-science-process/overview](https://learn.microsoft.com/es-es/azure/architecture/data-science-process/overview)*

*Massiris, M., Delrieux, C., & Fernández, J. Á. (2020). Detección de equipos de protección
personal mediante red neuronal convolucional Yolo. Actas de Las XXXIX Jornadas de
Automática, Badajoz, 5-7 de Septiembre de 2018.
<https://doi.org/10.17979/spudc.9788497497565.1022>*

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 122 de 128

ML: ventajas y desventajas de la regresión lineal – Barcelona Geeks. (2023, May 21). Retrieved from <https://barcelonageeks.com/ml-ventajas-y-desventajas-de-la-regresion-lineal>

Modelos de Detección de Objetos. (2020, December 19). Retrieved from <https://www.aprendemachinelearning.com/modelos-de-deteccion-de-objetos>

Nates-Parra, g. (2023). GENÉTICA DEL COMPORTAMIENTO:: ABEJAS COMO MODELO. Acta Biológica Colombiana, 16(3), 213–230. Retrieved from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-548X2011000300015.

Nelson, J. (2023). YoloV5 is Here: State-of-the-Art Object Detection at 140 FPS. Roboflow Blog. Retrieved from <https://blog.roboflow.com/YoloV5-is-here>

Olaya, M. (2017). Catálogo fotográfico de especies de flora apícola en los departamentos de Cauca, Huila y Bolívar. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt. https://repository.agrosavia.co/bitstream/handle/20.500.12324/13019/45044_60818.pdf?sequence=1&isAllowed=y

OpenAI. (2021). ChatGPT. Retrieved Month Day, Year, from <https://openai.com/research/chatgpt>

Pedro Arcos Méndez. (2019). Metodología KDD para Data Mining. Retrieved May 2, 2023, from Instituto Tecnológico Superior de los Ríos website: https://www.academia.edu/41660260/Metodolog%C3%ADa_KDD_para_Data_Mining

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 123 de 128

Pérez Porto, J., & Merino, M. (2021). *Entrenamiento - Qué es, definición y concepto*.

Definición.de. Retrieved from <https://definicion.de/entrenamiento>

pkranen. (2023, May 24). *Detección de objetos mediante Fast R-CNN - Cognitive Toolkit -*

CNTK. Retrieved from <https://learn.microsoft.com/es-es/cognitive-toolkit/object-detection-using-faster-r-cnn>

Procesamiento digital de imágenes - EcuRed. (2023, May 26). Retrieved from

https://www.ecured.cu/Procesamiento_digital_de_im%C3%A1genes

Qué es DataSet y para qué sirve | Data Science. (2023, January 14). Retrieved from [https://data-](https://data-science.es/que-es-dataset-y-para-que-sirve)

[science.es/que-es-dataset-y-para-que-sirve](https://data-science.es/que-es-dataset-y-para-que-sirve)

Redmon, J., & Farhadi, A. (2016). *Yolo9000: Better, Faster, Stronger*. Retrieved from

<https://arxiv.org/pdf/1612.08242.pdf>

Redmon, J., & Farhadi, A. (2018). *Yolov3: An Incremental Improvement*. Retrieved from

<https://arxiv.org/pdf/1804.02767.pdf>

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). *You Only Look Once: Unified, Real-*

Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern

Recognition (CVPR). <https://doi.org/10.1109/cvpr.2016.91>

Reza, Z. N. (2019). *Real-time Automated Weld Quality Analysis From Ultrasonic B-scan Using*

Deep Learning. Scholarship at UWindsor. Retrieved from

<https://scholar.uwindsor.ca/etd/7673>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 124 de 128

Rozada Raneros, S. (2021). Estudio de la arquitectura Yolo para la detección de objetos mediante deep learning. Retrieved from

<https://uvadoc.uva.es/bitstream/handle/10324/45359/TFM-G1316.pdf?sequence=1&isAllowed=y>

Saavedra, J. A. (2023). Regresión Lineal: qué es, para qué sirve, por qué es importante, tipos y ejemplos de uso. Ebac. Retrieved from <https://ebac.mx/blog/regreson-lineal>

Sandoval Serrano, L. J. (2018). Algoritmos de aprendizaje automático para análisis y predicción de datos. Revista Tecnológica; no.

11.http://redicces.org.sv/jspui/bitstream/10972/3626/1/Art6_RT2018.pdf

Sarmiento-Ramos, J. L. (2020). Aplicaciones de las redes neuronales y el deep learning a la ingeniería biomédica. Revista UIS Ingenierías, 19(4), 1–18. Retrieved from

<https://www.redalyc.org/journal/5537/553768213002/html>

Significados. (2016). Polinización. Significados. Retrieved from

<https://www.significados.com/polinizacion>

Sitio en mantenimiento. (2023). Retrieved May 7, 2023, from Wwww.gob.mx website:

<https://failover.www.gob.mx/mantenimiento.html>

Skalski, P. (2023, February 14). Make Sense. Retrieved from <https://skalskip.github.io/make-sense>

SkalskiP. (2023, May 23). make-sense. Retrieved from <https://github.com/SkalskiP/make-sense>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 125 de 128

Sojasingarayar, A. (2023). *Faster R-CNN vs Yolo vs SSD — Object Detection Algorithms.*

Medium. Retrieved from <https://medium.com/ibm-data-ai/faster-r-cnn-vs-Yolo-vs-ssd-object-detection-algorithms-18badb0e02dc>

Solawetz, J. (2023). *What is YoloV5? A Guide for Beginners.* Roboflow Blog. Retrieved from

<https://blog.roboflow.com/YoloV5-improvements-and-evaluation>

Srivastava, S., Divekar, A. V., Anilkumar, C., Naik, I., Kulkarni, V., & Patabiraman, V. (2021).

Comparative analysis of deep learning image detection algorithms. J. Big Data, 8(1), 1–27. doi: 10.1186/s40537-021-00434-w

SSD: *Understanding single shot object detection.* (2019, June 25). Retrieved from

<https://manalelaidouni.github.io/Single%20shot%20object%20detection.html>

Terven, J., & Cordova-Esparaza, D. (2023). *A COMPREHENSIVE REVIEW OF Yolo: FROM YoloV1 TO YoloV8 AND BEYOND UNDER REVIEW IN ACM COMPUTING SURVEYS.*

Retrieved from <https://arxiv.org/pdf/2304.00501.pdf>

Ultralytics. (2023, May 20). GitHub - ultralytics/YoloV5: YoloV5 🚀 in PyTorch > ONNX >

CoreML > TFLite. Retrieved May 20, 2023, from GitHub website:

<https://github.com/ultralytics/YoloV5>

Varroa destructor : USDA ARS. (2023, June 02). Retrieved from

<https://www.ars.usda.gov/northeast-area/beltsville-md-barc/beltsville-agricultural-research-center/bee-research-laboratory/docs/Varroa-destructor>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 126 de 128

Verastuchelova. (2023). *Destructor de Varroa*. Retrieved June 23, 2023, from Dreamstime

website: <https://es.dreamstime.com/photos-images/Varroa.html>

Villanueva, R. M. (2017). *Detección de objetos por computador*. Retrieved from

<https://uvadoc.uva.es/bitstream/handle/10324/45359/TFM-G1316.pdf?sequence=1&isAllowed=y>

Vista de Introducción al Aprendizaje Automático con Yolo. (2023). Retrieved May 21, 2023, from Ulatina.ac.cr website:

<https://revistas.ulatina.ac.cr/index.php/tecnologiavital/article/view/250/260>

Vista de Principales Metodologías en el Desarrollo de Proyectos de Minería de Datos. (2023).

Retrieved May 2, 2023, from Tecnocultura.org website:

<https://tecnocultura.org/index.php/Tecnocultura/article/view/9/9>

Wang, X., Wang, X., Hu, C., Dai, F., Xing, J., Wang, E., ... Guo, W. (2022). *Study on the Detection of Defoliation Effect of an Improved YoloV5x Cotton*. 12(10), 1583–1583.

<https://doi.org/10.3390/agriculture12101583>

Wang, X., Wang, X., Hu, C., Dai, F., Xing, J., Wang, E., ... Guo, W. (2022). *Study on the Detection of Defoliation Effect of an Improved YoloV5x Cotton*. 12(10), 1583–1583.

<https://doi.org/10.3390/agriculture12101583>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 127 de 128

YoloV5. (2022). Retrieved May 20, 2023, from Roboflow.com website:

<https://roboflow.com/model/Yolov7>

YoloV5. (2023). Retrieved May 20, 2023, from Roboflow.com website:

<https://roboflow.com/model/Yolov8>

Yolov6 Object Detection – Paper Explanation and Inference. (2022, October 11).

Retrieved May 22, 2023, from LearnOpenCV – Learn OpenCV, PyTorch, Keras, Tensorflow with examples and tutorials website: <https://learnopencv.com/Yolov6-object-detection/>

"¿Qué es el Proceso de ciencia de datos en equipo (TDSP)? - Azure Architecture Center.

(n.d.). Microsoft Learn. Retrieved May 16, 2023, from <https://learn.microsoft.com/es-es/azure/architecture/data-science-process/overview>

| ICA - Instituto Colombiano Agropecuario. (2022) En el Cauca, el ICA articula acciones para fortalecer la apicultura del departamento. Retrieved May 16, 2023, from Portal Corporativo ICA website: <https://www.ica.gov.co/noticias/ica-cauca-acciones-fortalecimiento-apicultura>

¿Explica brevemente las 3 principales desventajas de aplicar la metodología KDD en Machine Learning? (2023-05-21).

¿Qué es Deep Learning? | IBM. (2023, May 19). Retrieved from <https://www.ibm.com/es-es/topics/deep-learning>

¿Qué es la visión artificial? | IBM. (2023, May 25). Retrieved from <https://www.ibm.com/es-es/topics/computer-vision>

	DOCUMENTO FINAL PROYECTO DE GRADO	EDO – 01
		Versión 1
		Vigencia: 28/06/2023
		Página 128 de 128

¿Qué es la visión artificial? | Microsoft Azure. (2023, May 26). Retrieved from
<https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-computer-vision>

¿Qué es un dataset o un dataframe? - Talently. (2022, March 23). Retrieved from
<https://talently.tech/blog/que-es-un-dataset>

*“TÉCNICAS DE MACHINE LEARNING APLICADO A LA PREDICCIÓN DE FUGA DE
CLIENTES EN UNA INSTITUCIÓN DE SALUD PREVISIONAL BAJO LA
METODOLOGÍA KDD” DIEGO ALEJANDRO FLORES VELÁSQUEZ PROFESOR
GUÍA: DANILO GOMEZ CORREA MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL CONCEPCIÓN -CHILE DICIEMBRE, 2018. (n.d.).*

Retrieved from

https://repositorio.unab.cl/xmlui/bitstream/handle/ria/9012/a125556_Flores_D_Tecnicas_de_machine_learning_aplicado_2019.pdf?sequence=1&isAllowed=yhttps://rid.unrn.edu.ar/bitstream/20.500.12049/9096/3/Investigacion