# PROPUESTA DE MEJORAMIENTO DEL MODELO DE ML PARA EL SISTEMA DE MONITOREO Y SEGURIDAD LABORAL EN LA EMPRESA SLB

CAMILO ANDRES MELO TAYO


UNIVERSIDAD SANTO TOMÁS
DIVISIÓN DE INGENIERÍAS
FACULTAD DE INGENIERÍA DE TELECOMUNICACIONES
BOGOTÁ

# PROPUESTA DE MEJORAMIENTO DEL MODELO DE ML PARA EL SISTEMA DE MONITOREO Y SEGURIDAD LABORAL EN LA EMPRESA SLB

Trabajo de grado pasantías en la empresa SLB


Presenta:

CAMILO ANDRES MELO TAYO

Director:

PEDRO ALEJANDRO MANCERA LAGOS

INGENIERO ELECTRÓNICO

UNIVERSIDAD SANTO TOMÁS
DIVISIÓN DE INGENIERÍAS
FACULTAD DE INGENIERÍA DE TELECOMUNICACIONES


BOGOTÁ, 2024

Dedico este trabajo de grado con mucho amor a Dios y a mis padres quienes nunca se cansaron de soñar un futuro y una vida mejor para nosotros, quienes día a día me enseñan a ser mejor persona y me ayudan a levantarme siempre.
Espero algún día retribuirles todo lo que me han dado.

*"Con paciencia esperé que el SEÑOR me ayudara, y él se fijó en mí y oyó mi clamor. Me sacó del foso de desesperación, del lodo y del fango. Puso mis pies sobre suelo firme y a medida que yo caminaba, me estabilizó. Me dio un canto nuevo para entonar, un himno de alabanza a nuestro Dios. Muchos verán lo que él hizo y quedarán asombrados; pondrán su confianza en el SEÑOR. Ah, qué alegría para los que confían en el SEÑOR, los que no confían en los orgullosos ni en aquellos que rinden culto a ídolos."*
-Salmos 40:1-4 NTV

## AGRADECIMIENTOS

Quiero agradecer a Dios por la oportunidad que me da de poder soñar con un futuro mejor para mí y para mi familia, por siempre acompañarme en los momentos alegres y en los no tan alegres. Le doy gracias por no soltarme de su mano y por jamás rendirse conmigo hoy en día puedo ver lo que Él quería para mí y me llena el corazón de orgullo.

Agradezco a mi familia por su apoyo incondicional, a mis padres por nunca rendirse conmigo, porque día a día vivieron y sintieron lo conmigo lo que ellos no pudieron, mis viejos esta va por ustedes gracias por ser el motor de mi vida, y la inspiración por la cual todos los días me levanto. A mi hermana Erika por ser mi compañera de vida, mi confidente mi escudera, por siempre soñar con nosotros y “por ser la aguja que lleva el hilo” aquí va uno de esos resultados, ¡lo logramos! ustedes son los seres humanos más espectaculares de esta vida, los amo.

Agradezco a mi tutor por su tiempo dedicado durante el desarrollo de este trabajo y a mis profesores que por su esmero y dedicación durante este tiempo de formación hicieron que me esforzara y madurar en el camino de la vida, a mis compañeros que fueron un apoyo incondicional, gracias a todos por permitirme conocerlos y compartir un pedacito de mi vida con cada uno de ustedes, gracias a todas las personas que hicieron parte del proceso, cada uno de ustedes me dejan enseñanzas muy valiosas para mí como ser humano y como profesional.

# Tabla de contenido

## RESUMEN

Esta monografía surge como resultado de un trabajo de investigación realizado en el marco de las prácticas profesionales desempeñadas en la empresa SLB (Schlumberger) entre el 5 de Julio de 2023 y el 5 de enero de 2024, en el que fue posible identificar un problema en la implementación por parte de dicha empresa del sistema de monitoreo y seguridad laboral basado en inteligencia artificial (IA) implementado con el propósito de enviar alertas de seguridad respecto de las personas que estén manipulando en indebida forma las herramientas o elementos de seguridad suministrados por la compañía. La investigación se llevó a cabo *"in situ",* aprovechando el conocimiento adquirido durante la práctica profesional desarrollada dentro de la empresa, la cual, permitió identificar los desafíos específicos que enfrenta la compañía en la implementación del referido sistema. La propuesta presentada en este trabajo se centra en la optimización del sistema mencionado anteriormente, mediante la implementación de herramientas en la nube de Google Cloud Platform (GCP), para lo cual, se abordan los problemas identificados a través del análisis de soluciones y se plantea una propuesta que combina tecnologías de vanguardia como el modelo YOLO y técnicas de reentrenamiento con la base de datos COCO (Common Objects in Context) y data sets personalizados, junto con la infraestructura escalable y confiable de GCP.



*Palabras clave:* GCP; IA; Nube; CCTV; servidores.



## ABSTRACT

This monograph arises as a result of a research work carried out within the framework of the professional practices performed in the company SLB (Schlumberger) between July 5, 2023 and January 5, 2024, in which it was possible to identify a problem in the implementation by the company of the monitoring and work safety system based on artificial intelligence (AI) implemented for the purpose of sending safety alerts regarding people who are improperly manipulating the tools or safety elements supplied by the company. The research was carried out “in situ”, taking advantage of the knowledge acquired during the professional practice developed within the company, which allowed to identify the specific challenges faced by the company in the implementation of the referred system. The proposal presented in this work focuses on the optimization of the aforementioned system, through the implementation of tools in the Google Cloud Platform (GCP) cloud, for which, the problems identified through the analysis of solutions are addressed and a proposal that combines cutting-edge technologies such as the YOLO model and retraining techniques with the COCO (Common Objects in Context) database and customized data sets, along with the scalable and reliable infrastructure of GCP is proposed.



*Keywords:* GCP; AI; Cloud; CCTV; servers.

# INTRODUCCIÓN

La presente monografía, aborda la problemática actual en la implementación de modelos de inteligencia artificial (IA) específicamente machine learning (ML) para la seguridad laboral implementados por la empresa SLB (Schlumberger), líder en la industria petrolera. Se parte del contexto de la empresa, que ha destacado por su compromiso con la innovación, especialmente en la optimización de la exploración de pozos petroleros a través de IA. Sin embargo, la empresa presenta una brecha en el desarrollo de herramientas de IA a nivel interno destinadas a mejorar las condiciones de seguridad diarias de los empleados.

En ese sentido, el problema identificado en esta investigación, tiene que ver con la falta de eficiencia en la implementación de sistemas de IA por parte de SLB para monitorear a los trabajadores mediante sistemas de CCTV que permitan identificar el uso adecuado de elementos de protección y la correcta manipulación de herramientas de trabajo, con el fin de enviar alertas de seguridad respecto de los empleados que estén infringiendo las normas de seguridad, por ejemplo, el uso incorrecto de los elementos de protección o la manipulación incorrecta de las herramientas con las que trabajan. La referida falta de eficiencia está relacionada con la descarga y etiquetado manual de videos que conlleva a un gasto innecesario de recursos, pues la nube ofrece herramientas que permiten o facilitan automatizar este proceso.

Para solucionar el problema, la monografía plantea una solución para mejorar el sistema de monitoreo y seguridad laboral, optimizando procesos y mejorando la eficiencia operativa mediante una arquitectura en la nube que ofrece flexibilidad y escalabilidad, por lo que los servidores locales actuales requerirán una mayor capacidad de procesamiento y estarán tecnológicamente rezagados. Dicha situación, puede evitarse, si el sistema de monitoreo y seguridad laboral propuesto por la empresa se desarrolla en la nube, dados los beneficios que la nube ofrece al momento de implementar el modelo de IA en cada proyecto.

El objetivo general de esta investigación es presentar una propuesta para SLB que aborde los desafíos actuales en la implementación de sistema de monitoreo y seguridad laboral soportado en modelos de ML en sus zonas de trabajo alrededor del mundo. La solución

propuesta aprovecha herramientas avanzadas de IA y la infraestructura GCP para la propuesta de una arquitectura basada en estas herramientas.

A lo largo del documento, se detalla la metodología utilizada para desarrollar la propuesta, que incluye la identificación del problema, los datos obtenidos en la investigación realizada durante las prácticas profesionales en la empresa, y el diseño de una arquitectura en la nube utilizando servicios clave de IA y GCP. Finalmente, se presentan los resultados de entrenamiento del modelo de ML utilizado por SLB y el análisis de datos recolectados durante el tiempo de la práctica profesional, así como el diseño de un pipeline para la implementación y monitoreo continuo de la solución.

El valor agregado de esta investigación radica en su enfoque práctico y orientado a la resolución de problemas reales en el contexto empresarial actual. Se espera que los resultados y recomendaciones presentados en este documento sean de utilidad para SLB y otras empresas que enfrentan desafíos similares en la implementación de sistemas de IA para mejorar la seguridad laboral y la eficiencia operativa.

# 1 MARCO GENERAL

## 1.1 DESCRIPCIÓN DEL PROBLEMA

La empresa SLB, ha implementado un sistema de IA respaldado por ML, con el propósito de reconocer y clasificar personas, máquinas, herramientas y elementos de protección en puntos estratégicos de la organización. Estos puntos estratégicos están equipados con sistemas de CCTV que capturan eventos relacionados con la seguridad de los trabajadores, que sean relevantes o pongan en riesgo la seguridad de los mismos trabajadores a lo largo del día, y de esta manera el sistema de monitoreo y seguridad laboral implementado actualmente debe evaluar si las personas que transitan por dichas zonas actúan conforme a las instrucciones y el reglamento de seguridad diseñado por la empresa y de pasar lo contrario enviar alertas de seguridad a los administradores del sistema de monitoreo y seguridad laboral en SLB. Sin embargo, el proceso actual de entrenamiento e implementación de este sistema de monitoreo y seguridad laboral presenta deficiencias significativas que afectan la eficiencia y la efectividad del sistema.

El principal inconveniente radica en la búsqueda manual de eventos específicos relacionados con personas, máquinas y herramientas, para su posterior descarga y etiquetado manual para el uso posterior en el entrenamiento del modelo de ML. Este procedimiento implica una carga operativa considerable y limita la escalabilidad del sistema, ya que cada nueva ubicación que se incorpora requiere un reentrenamiento manual del modelo con nuevos videos que se enfocan en la ubicación en la cual será implementado. Además, la implementación actual carece de una estructura basada en la nube para respaldar tanto el modelo de inteligencia artificial como el almacenamiento de datos capturados por las cámaras en cada ubicación.

El reentrenamiento del modelo involucra la búsqueda y descarga de videos y la laboriosa tarea de etiquetar manualmente cada elemento que se espera que el modelo reconozca en

el futuro cuando se implemente. Esto incluye la identificación de personas sin elementos de protección adecuados y el uso indebido de herramientas, con el objetivo de activar alertas automáticas ante comportamientos de riesgo.

Otro aspecto crítico identificado es que el modelo se despliega localmente en cada ubicación, sin aprovechar las ventajas de la computación en la nube. Esta falta de integración en la nube limita la flexibilidad, la sincronización y la centralización de datos, afectando la capacidad de adaptarse eficientemente a nuevas ubicaciones y escenarios. Estas limitaciones afectan negativamente la eficiencia del sistema, la capacidad de escalar a nuevas ubicaciones y la optimización del proceso de reentrenamiento del modelo. En este contexto, se plantea la necesidad de desarrollar propuestas que aborden estas problemáticas y optimicen la implementación de la inteligencia artificial en el entorno laboral de SLB.

## 1.2 JUSTIFICACIÓN

En la actualidad, la implementación efectiva de tecnologías de IA en entornos empresariales es fundamental para mejorar la eficiencia operativa y mantener la competitividad en un mercado en constante evolución. La empresa SLB, consciente de ello, ha emprendido iniciativas para integrar modelos de IA en sus operaciones internas para optimizar procesos y mejorar la seguridad laboral de sus empleados.

Sin embargo, la implementación de modelos de IA en entornos empresariales multinacionales presenta desafíos significativos, especialmente en lo que respecta a la gestión de datos, el entrenamiento de modelos y el despliegue en múltiples ubicaciones. La falta de una infraestructura sólida en la nube y la ausencia de un enfoque integrado para el desarrollo y la implementación de modelos de IA pueden limitar el potencial de estos sistemas y obstaculizar su efectividad en el entorno laboral.

Por lo tanto, esta monografía pretende abordar estas limitaciones mediante el diseño de una propuesta de implementación de una arquitectura en la nube optimizada y un pipeline de desarrollo de IA robusto para SLB. El diseño de esta arquitectura en la nube se basa en servicios y tecnologías proporcionadas por GCP, aprovechando su escalabilidad, seguridad y capacidad de integración.

Esta propuesta consiste en un pipeline que abarque todas las etapas del ciclo de vida de desarrollo de modelos de IA, desde la captura de datos de las cámaras CCTV hasta el despliegue y la monitorización en tiempo real de los modelos en múltiples ubicaciones. Se emplearán servicios como Google AI Platform, Google Kubernetes Engine (GKE), Google Cloud Functions y otros, para automatizar y orquestar eficientemente el flujo de trabajo, proponiendo una gestión eficaz de los modelos de IA en toda la empresa.

Al proponer una infraestructura en la nube y el proceso de mejoramiento del sistema de reconocimiento basado en IA, SLB estará mejor posicionada para aprovechar el potencial de la inteligencia artificial en sus operaciones internas, mejorando la seguridad laboral, optimizando procesos y aumentando la eficiencia operativa. Esta monografía no solo propone una solución práctica para los desafíos actuales de SLB, sino que también sentará las bases para futuras innovaciones y mejoras en el uso de la inteligencia artificial en el entorno empresarial.

## 1.3 OBJETIVOS

### *1.3.1 OBJETIVO GENERAL*

Proponer una estrategia de despliegue basada en una arquitectura en la nube para el sistema de monitoreo y seguridad laboral mediante inteligencia artificial en Schlumberger (SLB), para optimizar los procesos internos en la empresa, permitiendo la automatización, escalabilidad y eficiencia en el entrenamiento y despliegue de sistemas de monitoreo y seguridad laboral basados en modelos de IA para reconocer personas y usar elementos de protección y herramientas, divididos por zonas en múltiples ubicaciones de la empresa.

### *1.3.2 OBJETIVOS ESPECÍFICOS*

- Describir el proceso actual de implementación del sistema de monitoreo y seguridad laboral en SLB.

- Identificar las ventajas y desventajas actuales con las que cuenta el sistema de monitoreo y seguridad laboral de SLB.

- Identificar las tecnologías y servicios disponibles para el despliegue de una arquitectura en la nube que facilite el proceso de implementación del sistema de monitoreo y seguridad laboral basado en ML en SLB.

- Analizar los beneficios del despliegue de una arquitectura en la nube en términos de eficiencia operativa, escalabilidad y seguridad, así como el impacto de la mejora de los modelos de ML en la precisión y generalización de los resultados.

- Proponer una estrategia para el despliegue de una arquitectura nube para la implementación de los sistemas de monitoreo y seguridad laboral en SLB.

## 1.4 ALCANCE

La propuesta de la monografía pretende implementar una arquitectura basada en los servicios de la nube, que optimice los procesos de implementación de IA de SLB para evitar procesos de etiquetado manual manual de imágenes para el entrenamiento de modelos de machine learning (ML) y tener mayor capacidad de procesamiento en el sistema de moniteoreo y seguridad laboral en SLB.

# 2 MARCO TEÓRICO

## 2.1 TECNOLOGÍAS Y SERVICIOS EN LA NUBE

En la era digital actual, las tecnologías y servicios en la nube han emergido como pilares fundamentales para el desarrollo y despliegue de aplicaciones avanzadas, incluyendo modelos de machine learning (ML). La nube ofrece una infraestructura robusta y flexible que permite a las empresas acceder a recursos computacionales de alta capacidad sin la necesidad de invertir en costosos equipos físicos. Se explorarán las principales plataformas y servicios que facilitan el despliegue de modelos de ML y cómo estos mejoran la eficiencia operativa, ofrecen escalabilidad y garantizan la seguridad en el contexto de aplicaciones de inteligencia artificial (IA).

Las plataformas de computación en la nube como Google GCP, Amazon Web Services (AWS) y Microsoft Azure proporcionan un ecosistema completo de servicios que apoyan el ciclo de vida de los modelos de machine learning. A continuación, se revisará como estas plataformas ofrecen servicios específicos para el almacenamiento, procesamiento, entrenamiento y despliegue de modelos de ML:

### *2.1.1 GCP*

GCP proporciona servicios como Google Compute Engine (GCE) para el entrenamiento de modelos en máquinas virtuales de alto rendimiento, Google AI Platform para el desarrollo, entrenamiento y despliegue de modelos de ML, y Google Kubernetes Engine (GKE) para la orquestación de contenedores y gestión de aplicaciones en contenedores. Además, GCP ofrece Google Cloud Storage (GCS) para el almacenamiento escalable de datos y Cloud Pub/Sub para la transmisión de datos en tiempo real [1][2][3][4].

### *2.1.2 Microsoft Azure*

Azure Machine Learning es la plataforma de Microsoft para el desarrollo y despliegue de modelos de ML, complementada por Azure VM's para la computación y Azure Blob Storage para el almacenamiento de datos. Azure Kubernetes Service (AKS) ofrece capacidades de orquestación de contenedores, mientras que Azure Event Hubs facilita la transmisión de datos en tiempo real [5].

### *2.1.3 AWS*

AWS ofrece servicios similares, como Amazon SageMaker para el desarrollo y entrenamiento de modelos de ML, Amazon EC2 para la computación elástica, y Amazon S3 para el almacenamiento de datos. AWS también proporciona herramientas de orquestación como Amazon EKS (Elastic Kubernetes Service) y servicios de transmisión de datos en tiempo real como Amazon Kinesis [6][7].

En este caso, de las tecnologías en la nube analizadas, se optará por las tecnologías y servicops ofrecidos por GCP, gracias a su gran variedad y adaptación que ofrece al implementar nuevas iniciativas como la que se abordará en este documento. El uso de tecnologías en la nube ofrece mejoras significativas en la eficiencia operativa, la automatización y orquestación de flujos de trabajo mediante la adopción de nuevas herramientas que hacen posible reducir el tiempo de desarrollo de nuevas solociones basadas en software, aumentando la productividad en los equipos de trabajo. Dos aspectos

clave en la mejora de la eficiencia operativa para los sistemas de monitoreo y seguridad laboral en SLB son:

### 2.1.4 Optimización de recursos

Los servicios en la nube permiten la asignación dinámica de recursos computacionales según la demanda, evitando el sobredimensionamiento y reduciendo costos. La capacidad de escalar recursos automáticamente según las necesidades del procesamiento de datos y el entrenamiento de modelos asegura que los recursos se utilicen de manera óptima.

### *2.1.5 Automatización de procesos*

Herramientas como Apache Airflow y Jenkins facilitan la orquestación y automatización de flujos de trabajo complejos, desde la ingesta de datos hasta el despliegue de modelos. Esto reduce la necesidad de intervención manual por parte de los desarrolladores y minimiza errores humanos, mejorando la consistencia y la confiabilidad del proceso de desarrollo e implementación de soluciones basadas en la nube. Por otro lado, herramientas como Terraform permite la gestión automatizada de la infraestructura, asegurando que los recursos se provisionen de manera consistente y eficiente [8][9][10].

En términos de escalabilidad, los servicios en la nube permiten ajustar los recursos según la demanda. Por ejemplo, GKE puede escalar automáticamente los contenedores que ejecutan los modelos de ML, garantizando que el sistema pueda manejar cargas de trabajo variables sin interrupciones. Este escalado automático es crucial para aplicaciones que deben procesar datos en tiempo real o durante picos de actividad [3].

La seguridad es otro aspecto crítico garantizado por las tecnologías en la nube, permiten la gestión segura de secretos y datos sensibles, mientras cifran las comunicaciones para proteger los datos en tránsito. Además, la configuración de Identity and Access Management (IAM) y el uso de Cloud Identity-Aware Proxy aseguran que solo usuarios autorizados puedan acceder a los servicios y datos críticos. Estas medidas de seguridad son fundamentales para proteger la integridad y confidencialidad de los datos en aplicaciones de IA [11][12].

## 2.2 INTRODUCCIÓN AL MACHINE LEARNING

El aprendizaje automático (machine learning) es una rama de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos que permiten a las computadoras aprender, hacer predicciones o tomar decisiones basadas en datos. A diferencia de los sistemas tradicionales de programación, donde las reglas y lógica son explícitamente codificadas, los sistemas de machine learning se entrenan con grandes volúmenes de datos para identificar patrones y hacer inferencias. Esta capacidad de aprendizaje y adaptación es lo que hace que el machine learning sea una herramienta poderosa para una amplia variedad de aplicaciones, desde la visión por computadora hasta el procesamiento del lenguaje natural.

A medida que las empresas buscan mejorar sus procesos y tomar decisiones más informadas, el ML se ha convertido en una herramienta indispensable. Este apartado proporciona una visión general de los fundamentos del aprendizaje automático y explica los diferentes tipos de aprendizaje automático destacando su relevancia en el contexto del proyecto.

### *2.2.1 Fundamentos del aprendizaje automático*

En términos generales, el proceso de machine learning implica tres etapas principales: la recopilación y preparación de datos, el entrenamiento del modelo, y la evaluación y despliegue del modelo. La recopilación y preparación de datos es crucial, ya que la calidad y cantidad de datos disponibles tienen un impacto directo en el rendimiento del modelo. Una vez que los datos están preparados, el siguiente paso es seleccionar y entrenar un modelo. Este proceso implica elegir un algoritmo adecuado y ajustar sus parámetros utilizando un conjunto de datos de entrenamiento. Finalmente, el modelo se evalúa con un conjunto de datos de prueba para verificar su precisión y capacidad de generalización [13].

El aprendizaje automático se basa en la idea de que los sistemas pueden identificar patrones y relaciones en los datos y utilizar esta información para realizar tareas específicas sin ser explícitamente programados para cada una. Los modelos de machine learning se entrenan con datos históricos para predecir o clasificar nuevas observaciones. Este proceso implica la selección de un modelo adecuado, el entrenamiento del modelo con datos etiquetados (en el caso del aprendizaje supervisado), la evaluación del rendimiento del modelo y la implementación del modelo entrenado en un entorno de producción [13].

### *2.2.2 Tipos de aprendizaje automático*

- **Aprendizaje supervisado:** Aprendizaje supervisado: En este tipo de aprendizaje, el modelo se entrena con un conjunto de datos etiquetados, donde cada ejemplo de entrenamiento incluye una entrada y la salida deseada. Los algoritmos de aprendizaje supervisado intentan aprender la relación entre las entradas y las salidas para poder predecir la etiqueta de nuevas entradas. Este tipo de aprendizaje es útil en tareas como la clasificación de imágenes, el reconocimiento de voz y la predicción de tendencias de mercado. En el contexto del proyecto, el aprendizaje

supervisado es crucial para el desarrollo de modelos que puedan detectar y reconocer personas, cascos, guantes y herramientas en las grabaciones de cámaras CCTV [14].

- **Aprendizaje no supervisado:** A diferencia del aprendizaje supervisado, en el aprendizaje no supervisado, los datos de entrenamiento no están etiquetados, el objetivo es encontrar estructuras ocultas o patrones en los datos. Los algoritmos no supervisados son útiles para tareas como el clustering (agrupamiento de datos similares) y la reducción de dimensionalidad (simplificación de datos complejos). En el contexto del proyecto, el aprendizaje no supervisado podría aplicarse para analizar patrones de comportamiento en los datos de video sin necesidad de etiquetas explícitas [15].

- **Aprendizaje por refuerzo:** En este enfoque, un agente (modelo de ML) aprende a tomar super-decisiones secuenciales interactuando con su entorno. El agente recibe recompensas o penalizaciones basadas en sus acciones y ajusta su estrategia para maximizar la recompensa acumulada. El aprendizaje por refuerzo es especialmente útil en situaciones donde la toma de decisiones debe adaptarse a un entorno dinámico y en tiempo real. Aunque menos común en el análisis de video, este tipo de aprendizaje puede ser relevante en la optimización de procesos y la mejora continua en el sistema de monitoreo y seguridad laboral [16].

### *2.2.3 Relevancia en el proyecto*

En el contexto del proyecto para el sistema de monitoreo y seguridad laboral, el aprendizaje supervisado es particularmente relevante. Por ejemplo, los modelos de detección de objetos como YOLO se entrenan utilizando conjuntos de datos etiquetados para reconocer y clasificar objetos específicos en imágenes y videos [14]. Estos modelos pueden identificar

personas, cascos, guantes y herramientas, lo que es crucial para mejorar la seguridad en el entorno laboral.

El aprendizaje no supervisado también puede ser útil en la fase de exploración de datos, ayudando a descubrir patrones y anomalías que no son evidentes a simple vista [15]. Por último, aunque el aprendizaje por refuerzo no se aplica directamente en este proyecto, su capacidad para optimizar decisiones a través de la retroalimentación constante podría ser investigada para futuras mejoras en la toma de decisiones automatizada [16].

## 2.3 MODELOS DE DETECCIÓN DE OBJETOS

Los modelos de detección de objetos son una clase de algoritmos de ML que identifican y localizan objetos en imágenes y videos, estos modelos son esenciales para una variedad de aplicaciones, incluyendo seguridad, automatización industrial, vehículos autónomos y más. La detección de objetos no solo implica clasificar un objeto dentro de una imagen, sino también determinar su posición mediante el uso de bounding boxes (cuadro delimitador) o regiones.

### *2.3.1 Arquitecturas y modelos de ML más usados en la detección de objetos*

- **YOLO (You Only Look Once):** YOLO es un algoritmo de detección de objetos en tiempo real que destaca por su rapidez y eficiencia. A diferencia de otros enfoques, YOLO divide una imagen en una cuadrícula y simultáneamente predice bounding boxes y probabilidades de clase para cada celda en una sola pasada. Este método permite que YOLO sea extremadamente rápido, procesando imágenes en tiempo real, lo que lo hace adecuado para aplicaciones que requieren velocidad, como la vigilancia en video y la navegación autónoma. YOLOv10, una de las versiones más

recientes, introduce optimizaciones que mejoran tanto la precisión como la velocidad del modelo, utilizando técnicas avanzadas de anclaje y procesamiento de imágenes [17][18].

- **SSD (Single Shot MultiBox Detector):** SSD es otro algoritmo de una sola etapa que, al igual que YOLO, predice bounding boxes y probabilidades de clase directamente desde la imagen completa en una sola pasada, sin embargo, SSD utiliza múltiples mapas de características de diferentes resoluciones, lo que le permite detectar objetos de varios tamaños de manera más efectiva. Este enfoque permite a SSD mantener un equilibrio favorable entre velocidad y precisión, lo que lo hace ideal para aplicaciones móviles y embebidas donde los recursos de procesamiento son limitados. SSD utiliza múltiples mapas de características en diferentes escalas para detectar objetos de varios tamaños, lo que contribuye a su alta precisión en la detección de objetos pequeños y grandes [19].

- **R-CNN (Region-based Convolutional Neural Networks):** La familia de modelos R-CNN, que incluye variantes como fast R-CNN y faster R-CNN, emplea un enfoque de dos etapas para la detección de objetos. Primero, generan propuestas de regiones que podrían contener objetos utilizando métodos selectivos de búsqueda o redes específicas para la propuesta de regiones. Luego, cada propuesta de región es clasificada y refinada por una CNN, aunque los modelos R-CNN son conocidos por su alta precisión, su enfoque de dos etapas puede ser más lento en comparación con otros algoritmos. Fast R-CNN y faster R-CNN han mejorado la velocidad al optimizar la extracción de características y la propuesta de regiones, respectivamente, pero siguen siendo menos eficientes que modelos de una sola etapa como YOLO y SSD [20].

- **RetinaNet:** RetinaNet introduce una innovación significativa con la Focal Loss, una función de pérdida diseñada para manejar el problema de los datos

desbalanceados, donde las clases de fondo son mucho más frecuentes que las clases de objeto. Esta función de pérdida ajusta el peso de las clases de fondo, permitiendo que el modelo se enfoque más en las clases de objeto. RetinaNet mantiene un alto nivel de precisión y es particularmente eficaz en la detección de objetos pequeños y difíciles de identificar, gracias a su estructura de red piramidal que captura características a diferentes niveles de resolución. Esto lo hace especialmente útil en aplicaciones donde la precisión es crucial, como en la identificación de anomalías en imágenes médicas o la vigilancia detallada [21].

### *2.3.2 Aplicaciones en los modelos de detección de objetos*

La detección de objetos es una tecnología avanzada de machine learning que tiene una amplia gama de aplicaciones en diferentes sectores, utilizando algoritmos y modelos sofisticados, esta tecnología puede identificar y rastrear objetos específicos dentro de imágenes y videos, proporcionando observaciones valiosas y mejorando la eficiencia operativa en diversas áreas. A continuación, se exploran algunas de las principales aplicaciones de la detección de objetos:

- **Seguridad y vigilancia:** modelos de detección de objetos son fundamentales en los sistemas de seguridad y videovigilancia. Estas tecnologías permiten identificar y rastrear personas y objetos en tiempo real, mejorando significativamente la seguridad en áreas públicas y privadas. Al detectar comportamientos inusuales o la presencia de objetos sospechosos, los sistemas de videovigilancia equipados con detección de objetos pueden alertar a los operadores de seguridad para una respuesta inmediata. Además, estas soluciones pueden integrarse con sistemas de análisis de video para proporcionar una supervisión automatizada, reduciendo la carga de trabajo manual y aumentando la precisión en la detección de amenazas potenciales [22].

- **Automóviles autónomos:** En el ámbito de los vehículos autónomos, la detección de objetos es una tecnología crítica para la seguridad y la navegación. Los modelos de detección de objetos permiten a los vehículos autónomos identificar peatones, otros vehículos, señales de tráfico y obstáculos en la carretera. Esta capacidad es esencial para la toma de decisiones en tiempo real, como frenar, acelerar o cambiar de carril, garantizando un manejo seguro y eficiente. La detección de objetos también contribuye a la creación de mapas precisos del entorno, mejorando la capacidad del vehículo para adaptarse a diferentes condiciones de conducción y responder adecuadamente a situaciones imprevistas [23].

- **Salud y medicina:** En el campo de la salud y la medicina, los modelos de detección de objetos han revolucionado el diagnóstico y tratamiento de diversas enfermedades. En la radiología, por ejemplo, estos modelos pueden ayudar a identificar tumores, fracturas y otras anomalías en las imágenes médicas con alta precisión. Esto no solo mejora la rapidez y la exactitud del diagnóstico, sino que también puede detectar condiciones que podrían pasarse por alto en los análisis manuales. Al automatizar el proceso de detección, los profesionales de la salud pueden enfocarse en la interpretación de los resultados y en la planificación del tratamiento, mejorando así los resultados para los pacientes [24].

- **Retail y comercio:** En el sector retail, la detección de objetos ofrece soluciones innovadoras para el análisis de comportamiento de los clientes, la gestión de inventarios y la prevención de pérdidas. En las tiendas minoristas, los modelos de detección de objetos pueden analizar patrones de movimiento y comportamiento de los clientes, proporcionando datos valiosos para optimizar la disposición de los productos y mejorar la experiencia de compra. Además, la detección automática de productos en los estantes puede ayudar en la gestión de inventarios, asegurando que los productos estén siempre disponibles y correctamente ubicados. Asimismo, estos sistemas pueden identificar y prevenir actividades de robo, mejorando la seguridad y reduciendo las pérdidas en el comercio minorista [25].

- **Agricultura:** En la agricultura, la detección de objetos puede utilizarse para monitorear la salud de los cultivos, identificar plagas y enfermedades, y optimizar el uso de recursos como el agua y los fertilizantes. Los drones equipados con cámaras y algoritmos de detección de objetos pueden sobrevolar campos y proporcionar datos en tiempo real sobre el estado de las plantas, permitiendo a los agricultores tomar decisiones informadas y mejorar la productividad. También son utilizados para el control de calidad y separación de algunos alimentos mediante el uso de computer vision que da órdenes a las maquinas industriales para separar los alimentos que cumplen los estándares de calidad de los que no los cumplen, asegurando o elevando el porcentaje de calidad en la selección de la mejor materia prima para cada empresa [26].

- **Logística y almacenamiento:** En el sector de la logística y el almacenamiento, los modelos de detección de objetos pueden mejorar la eficiencia y precisión del manejo de inventarios, estos sistemas pueden rastrear y gestionar el movimiento de bienes dentro de almacenes, optimizar rutas de picking y asegurar que los productos se almacenen y envíen correctamente. Esto reduce errores y aumenta la eficiencia operativa, lo que es crucial en centros de distribución de gran escala [27].

- **Realidad aumentada (AR) y realidad virtual (VR):** En aplicaciones de AR y VR, la detección de objetos es fundamental para la interacción y la experiencia del usuario. Por ejemplo, en AR, los modelos de detección de objetos pueden identificar superficies y objetos en el entorno del usuario, permitiendo la superposición de información digital y la interacción con elementos virtuales en el mundo real. Esto es útil en aplicaciones educativas, juegos, diseño y muchas otras áreas [28].

### *2.3.3 Limitaciones en los modelos de detección de objetos*

Aunque los modelos de detección de objetos han avanzado significativamente y se utilizan en una amplia variedad de aplicaciones, presentan varias limitaciones que afectan su rendimiento y usabilidad. A continuación, se describen algunos de los principales desafíos que enfrentan estos modelos:

- **Desafíos en tiempo real:** Modelos como YOLO y SSD conocidos por su capacidad para realizar detecciones rápidamente, lo que los hace adecuados para aplicaciones en tiempo real, sin embargo, a pesar de su velocidad, aún pueden enfrentarse a desafíos significativos en escenarios con alta demanda computacional. Aplicaciones como la videovigilancia en tiempo real o los vehículos autónomos requieren una capacidad de procesamiento constante y robusta, lo que significa que la latencia mínima es crítica a la hora de tomar decisiones en estas soluciones, y cualquier retraso puede tener consecuencias graves. Además, mantener un rendimiento constante a alta velocidad puede ser difícil en hardware con limitaciones de recursos, lo que obliga a optimizar continuamente los algoritmos y el hardware [29].

- **Escalabilidad:** La escalabilidad es otro desafío importante para los modelos de detección de objetos dado que la precisión de estos modelos puede disminuir cuando se aplican a conjuntos de datos muy grandes y diversos. Esto es porque los modelos deben generalizar bien a varias condiciones y escenarios, lo que es difícil de lograr, además, la necesidad de etiquetar manualmente grandes volúmenes de datos para el entrenamiento es un obstáculo significativo. El proceso de etiquetado es laborioso y costoso, y cualquier error en el etiquetado puede afectar la precisión del modelo. Soluciones como el aprendizaje automático semi-supervisado y no supervisado son alternativas que pueden ser contempladas en un largo plazo, pero todavía enfrentan desafíos en términos de precisión y aplicabilidad práctica [30].

- **Objetos pequeños y ocultos:** La detección de objetos pequeños o parcialmente ocultos sigue siendo un desafío considerable para muchos modelos, pues estos modelos pueden tener dificultades para identificar correctamente objetos que están en el fondo de una imagen o video o parcialmente cubiertos por otros objetos, o que son muy pequeños comparados con el resto de la imagen. Esto resulta siendo un limitante en su efectividad en escenarios complejos donde es crucial detectar todos los objetos presentes, técnicas como la utilización de redes de mayor resolución y el enfoque en características de múltiples escalas pueden ayudar, pero no resuelven completamente el problema [31].

- **Ambientes cambiantes:** Los modelos de detección de objetos pueden tener dificultades en condiciones adversas, como iluminación pobre, obstrucciones, la variabilidad en las apariencias de los objetos, el ángulo de la cámara y el entorno pueden afectar negativamente la precisión de los modelos de detección de objetos, estos modelos suelen ser entrenados en condiciones controladas y pueden tener dificultades para adaptarse a variaciones inesperadas en el entorno. Por ejemplo, una cámara de seguridad puede funcionar bien durante el día, pero su rendimiento puede desmejorar significativamente en condiciones de poca luz o cuando hay sombras fuertes. De manera similar, obstrucciones parciales pueden llevar a falsos negativos, donde el modelo no detecta un objeto presente [31].

- **Datos desbalanceados:** Los modelos de detección de objetos suelen enfrentar el problema de datos desbalanceados, donde algunas clases de objetos son más comunes que otras, lo que puede llevar a una detección desigual y a que el modelo rinda inferior al identificar las menos representadas. Por ejemplo, en un sistema de vigilancia, podría haber muchas más imágenes de personas que de objetos específicos como guantes o cascos, lo que lleva al modelo a ser menos efectivo en detectar estos últimos. Para abordar este problema, se pueden emplear técnicas

avanzadas como la Focal Loss, que ajusta la función de pérdida para poner más énfasis en las clases minoritarias, mejorando así el desempeño general del modelo [31].

## 2.4 TÉCNICAS DE ENTRENAMIENTO EN MODELOS DE MACHINE LEARNING

El entrenamiento de modelos de ML, en particular los destinados a la detección de objetos, requiere una combinación de métodos y estrategias específicas para asegurar la precisión y la eficiencia del modelo. A continuación, se detallan algunas de las técnicas más relevantes y las mejores prácticas en este campo.

### *2.4.1 Métodos y Estrategias de Entrenamiento*

- **División del conjunto de datos:** El primer paso en el entrenamiento de un modelo es dividir el conjunto de datos en tres partes: entrenamiento, validación y prueba. El conjunto de entrenamiento se utiliza para ajustar los parámetros del modelo, el conjunto de validación se emplea para sintonizar hiperparámetros y evitar el sobreajuste, y el conjunto de prueba se reserva para evaluar el desempeño final del modelo [32].

- **Aumentación de datos (Data Augmentation):** Para mejorar la robustez del modelo y reducir el riesgo de sobreajuste, se emplean técnicas de aumentación de datos, que generan nuevas muestras de datos a partir de las existentes mediante transformaciones como rotaciones, cambios de escala, traslaciones y ajustes de brillo o contraste. Albumentation es una biblioteca popular que proporciona una amplia gama de técnicas de aumentación de imágenes, permitiendo la generación de variaciones del conjunto de datos original [33] [34].

- **Transfer Learning:** Este método consiste en aprovechar modelos pre-entrenados en grandes data sets, como COCO, y adaptarlos a un nuevo problema específico. Transfer learning es especialmente útil cuando se dispone de un conjunto de datos limitado, ya que permite reutilizar conocimientos adquiridos por el modelo en tareas anteriores [35].

- **Fine-Tuning:** Similar a transfer learning, el fine-tuning implica ajustar un modelo pre-entrenado en un nuevo conjunto de datos, sin embargo, en lugar de entrenar solo las capas finales, se reentrena una parte o toda la red con un learning rate más bajo para adaptar el modelo a nuevas clases o características [36].

### *2.4.2 Mejores Prácticas y Desafíos*

Las mejores prácticas son esenciales en cualquier campo de trabajo, y en el contexto del desarrollo y despliegue de modelos de machine learning y sistemas de detección de objetos, su importancia se destaca en varios aspectos clave:

- **Preprocesamiento de datos:** Limpiar y normalizar los datos antes del entrenamiento es esencial para garantizar que el modelo aprenda de manera efectiva. Esto incluye la eliminación de datos duplicados o erróneos y la normalización de los valores de píxeles en las imágenes.

- **División del conjunto de datos:** Es crucial dividir el conjunto de datos en conjuntos de entrenamiento, validación y prueba. Esto asegura que el modelo no solo aprende

de los datos, sino que también se valida su capacidad de generalización en datos no vistos.

- **Entrenamiento en Lotes (Batch Training):** Dividir el conjunto de datos de entrenamiento en pequeños lotes y actualizar los parámetros del modelo después de cada lote ayuda a estabilizar el proceso de aprendizaje y a aprovechar mejor la memoria disponible.

- **Batch normalization:** Normalizar las entradas de cada capa para tener una media y desviación estándar estables durante el entrenamiento puede acelerar la convergencia y mejorar la precisión del modelo.

- **Validación cruzada (Cross-Validation):** Implementar técnicas de validación cruzada, como k-fold, ayuda a obtener una evaluación más robusta del desempeño del modelo y a reducir el riesgo de sobreajuste.

- **Regularización:** Para prevenir el sobreajuste (overfitting), se emplean técnicas de regularización como el dropout, L2 regularization y early stopping. Estas técnicas ayudan a que el modelo generalice mejor al no depender excesivamente de los datos de entrenamiento.

Al tratar grandes cantidades de datos o implementar sistemas robustos en los que se integre el ML y la IA se pueden abordar grandes desafíos durante el desarrollo e implementación de estos sistemas, pero enfrentar estos desafíos es fundamental porque fomenta la innovación, promueve la mejora continua y ayuda a desarrollar habilidades y conocimiento en el campo del ML y la detección de objetos. Algunos de estos desafíos son:

- **Desbalanceo de clases:** En la detección de objetos, a menudo hay una distribución desbalanceada de clases, donde algunas clases de objetos aparecen mucho más frecuentemente que otras. Esto puede llevar a un modelo que no generaliza bien.

- **Tiempo y Recursos Computacionales:** El entrenamiento de modelos complejos como YOLO puede ser intensivo en términos de tiempo y recursos computacionales. Utilizar hardware especializado como GPU's y TPU's, y servicios en la nube como GCE puede acelerar significativamente este proceso.

- **Ajuste de hiperparámetros:** Encontrar la configuración óptima de hiperparámetros (como la tasa de aprendizaje, el tamaño del lote y la arquitectura del modelo) es un proceso tedioso y crítico. Es por esto por lo que herramientas automatizadas como Google AI Platform, pueden facilitar este proceso de desarrollo e implementación de modelos de ML.

# 3 SITUACIÓN ACTUAL DE LA IMPLEMENTACIÓN EL SISTEMA DE MONITOREO Y SEGURIDAD LABORAL EN LA EMPRESA SLB

El proceso de implementación y montaje del modelo de Inteligencia artificial se realiza en equipos y servidores locales para cada ubicación geográfica de la empresa donde se va a poner en marcha este servicio, cada ubicación cuenta con al menos 15 cámaras con las que el modelo debe reentrenarse con un data-set hecho desde cero dado que el ángulo y la ubicación de cada cámara cambia y produce fallos en el modelo afectando los porcentajes de confidencialidad con los que el modelo reconoce a los trabajadores y las herramientas. Una vez ubicada cada cámara en el lugar donde va a trabajar en conjunto con el modelo, una persona se encarga de revisar las grabaciones del periodo de tiempo que grabo la cámara antes de implementar el modelo y realiza cortes de video donde ocurren los eventos que se necesitan para entrenar al modelo con el fin de que en un futuro este reconozca el mismo actuar de las personas de manera independiente, estos fragmentos de video se recortan y se descargan para luego ser divididos por frames para realizar el proceso de etiquetado de imágenes y entrenamiento del modelo.

## 3.1 ETIQUETADO DE IMÁGENES

En cada ubicación, se realiza el proceso de etiquetado de imágenes con la herramienta LabelImg. Se toman frames de video en los cuales aparecen los objetos que se desean detectar, que incluyen personas, cascos, guantes y herramientas. Cada objeto recibe su etiqueta correspondiente y se asigna a una clase especifica dentro del programa que servirá después para identificar con qué objeto se entrenará al modelo.

El proceso de corte y descargar de videos se hace a partir de las cámaras ubicadas en diferentes locaciones del mundo, el propósito de este proceso es obtener información útil y

especifica, para luego extraer o fragmentar dicha información en imágenes para su posterior etiquetado pudiendo así alimentar la generación de modelos de detección de objetos necesarios.

El proceso de corte y descarga se describe a continuación:

Una vez asignadas las cámaras y la estación de trabajo o Workstation, lo primero que se hace es conectarse a dicha estación de trabajo mediante conexión remota para hacer uso del software Ocularis en donde se puede acceder a las cámaras de las distintas ubicaciones de la empresa a nivel global. Una vez seleccionado el servidor de la ubicación en donde se va a trabajar como se logra evidenciar en la figura 1, se podrá acceder a todas las cámaras disponibles en una lista evidenciada en la figura 2, de esa lista se selecciona la cámara de interés o asignada, para que aparezca el video en tiempo real de la cámara seleccionada con anterioridad.

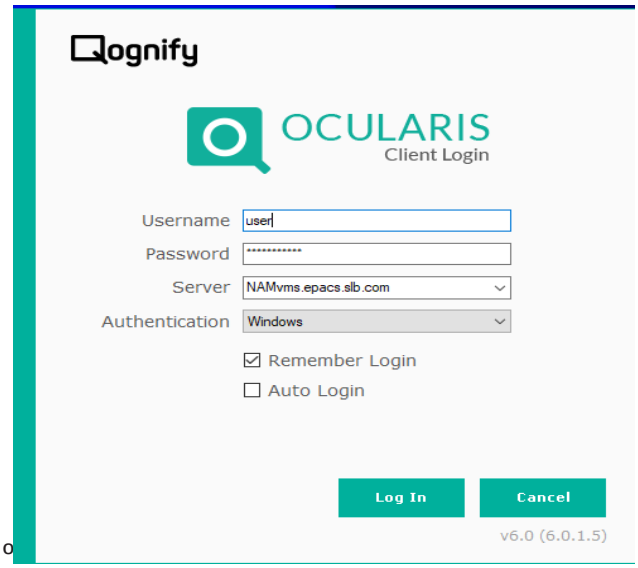


Figura 1. Interfaz de acceso para plataforma Ocularis

*Fuente: Ocularis*

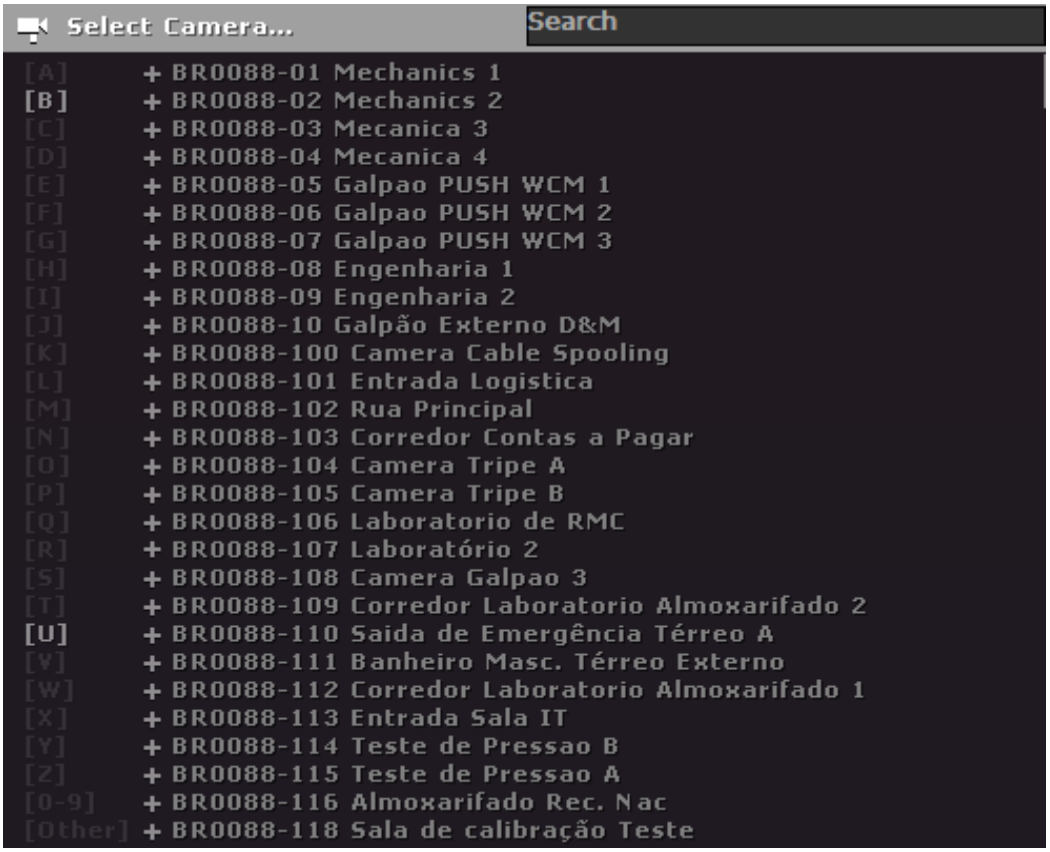


Figura 2. Lista de zonas, según la ubicación del servidor al que se accedió

*Fuente: Ocularis*

Para poder acceder a las cámaras en tiempo real y a sus grabaciones previas hay un tiempo de espera entre 4- 20 minutos, para revisar lo que han grabado se puede acceder con la opción ‘Browse’ que permite ir hasta 1 mes atrás en la grabación.

Se debe tener en cuenta el tipo de información que se va a descargar para el entrenamiento del modelo si es ‘mechanical lifting’ o ‘zone intrusions’. Mechanical lifting se refiere al movimiento de la herramienta con las que trabaja la empresa, de esta manera lo que se

busca es extraer videos de todo movimiento de dicha herramienta durante el lapso que se pueda observar en la cámara, este proceso se culmina una vez completa la meta de descargas para el entrenamiento del modelo, generalmente es de 100 eventos por zona.

Por otro lado, ‘zone intrusions’ se refiere a los momentos en que se puedan observar en la cámara indicada a operarios o trabajadores estando dentro de un área de trabajo restringida mientras la maquinaria se encuentra en funcionamiento, y la meta es la misma. Una vez obtenidos los satos se exportan en formato de base de datos como se muestra en la figura 3, con el formato de guardado de archivo que indiquen, generalmente se exporta a data base format, que se guarda en formato CSV (Comma Separated Value).

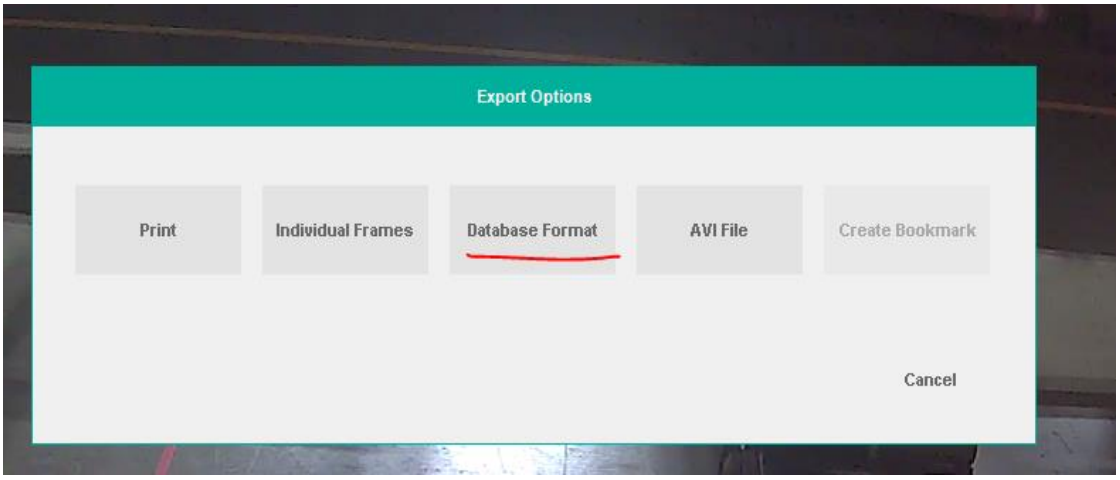


Figura 3. Interfaz de exportación de recorte de grabaciones.

*Fuente: Ocularis*

## 3.2 CREACIÓN DE DATAFRAMES

Dependiendo del objetivo con el que requieran entrenar el modelo, se utiliza labelImg o Labelme, para la primer herramienta dentro de los procesos de SLB se utiliza generalmente para la creación de etiquetas “simples” como lo son el etiquetado de personas, cascos y guantes como se puede observar en la figura 4 estas etiquetas se separan por clases o tipos de etiqueta para que el modelo reconozca que tipo de etiqueta se le está dando para que se entrene, estas clases se configuran y se guardan en un archivo de texto para luego ser seleccionadas en la creación de etiquetas, el objeto que se vaya a etiquetar se encierra dentro de un caja recta (Create RectBox) como se ve en el menú de la parte izquierda de la figura 4 y se abre una ventana emergente que da las opciones de nombre de etiquetas que se pueden seleccionar que fueron guardadas en el archivo de texto mencionado anteriormente, o también se pueden crear de manera directa en esta ventana en la parte superior simplemente introduciendo el nombre de la nueva etiqueta como se evidencia en la figura 5, en algunas ocasiones se solicita etiquetar herramientas rectas sin formas poligonales usadas para procesos de fracking y se realiza el mismo proceso descrito anteriormente.

Además, esta herramienta también es usada para rectificar el proceso de etiquetado realizado por otros trabajadores y corregir algunos errores que pueda haber dentro del conjunto de datos, a esto se le llama aprendizaje supervisado y es para reducir lo máximo posible el número de errores de datos que se puedan presentar para el modelo en el proceso de ejecución del reaprendizaje para el modelo de IA.

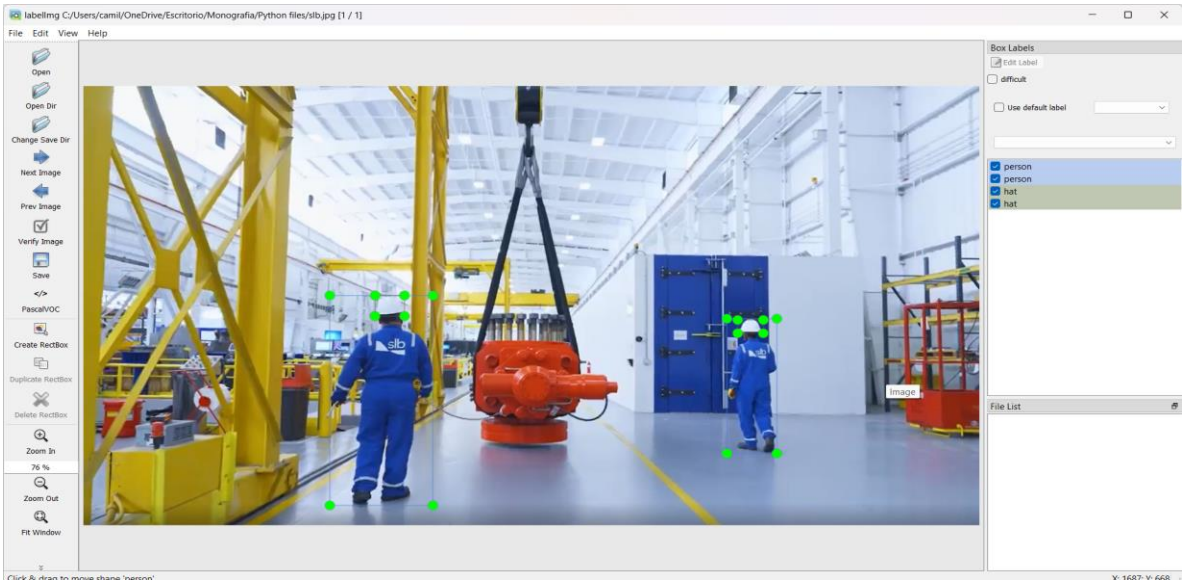


Figura 4. Interfaz de LabelImg.

*Fuente: LabelImg*

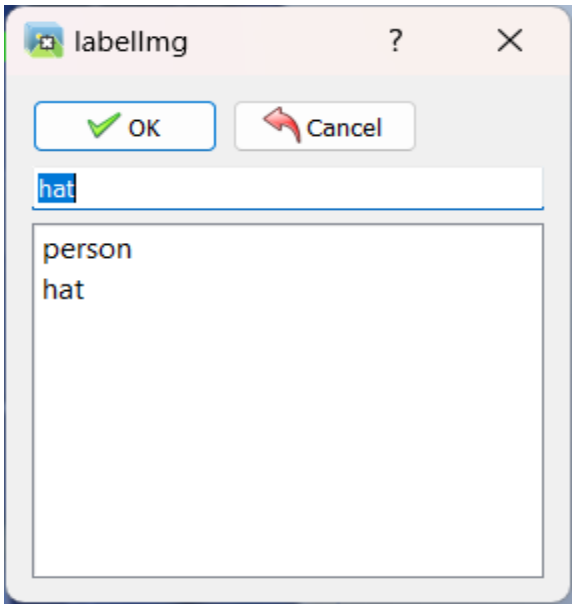


Figura 5. Ventana emergente para la asignación de tipo de etiqueta.

*Fuente: LabelImg*

Por otro lado se cuenta con Labelme, esta herramienta dentro de SLB únicamente es usada para el proceso de etiquetado de herramientas con bordes poligonales, puesto que encerrarlas en un rectángulo o cuadrado no es muy óptimo al momento de enseñar al modelo a reconocer estas herramientas, dada la gran variedad de herramientas con las que cuenta la empresa que pueden llegar a parecerse una entre y otra, el modelo puede llegar a ocasionar errores en el reconocimiento si no se realiza este proceso de manera adecuada, es por esto que la mejor opción es Labelme, para delimitar los límites de la herramienta que se dese reconocer mediante IA y enseñarle al modelo de IA cual herramienta es correcta y debe reconocer. El proceso de etiquetado es exactamente el mismo al anterior de LabelImg con la diferencia que en el menú de la interfaz no dice “create RectBox” sino "create Polygon”, como se evidencia en la figura 6.

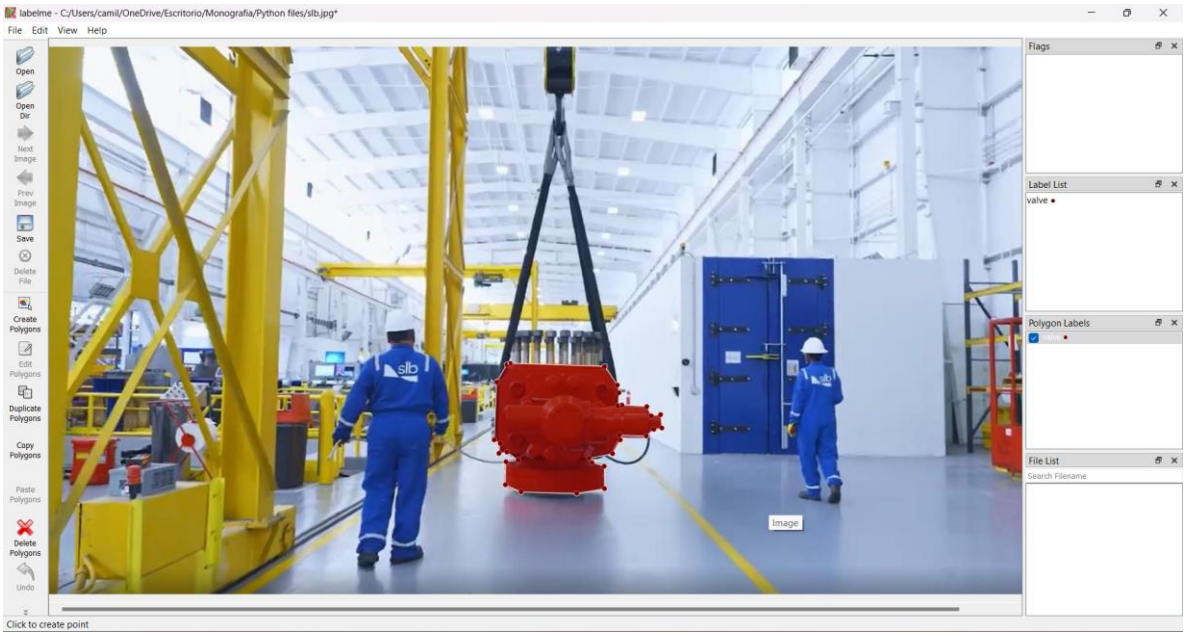


Figura 6. Interfaz de Labelme.

*Fuente: Labelme*

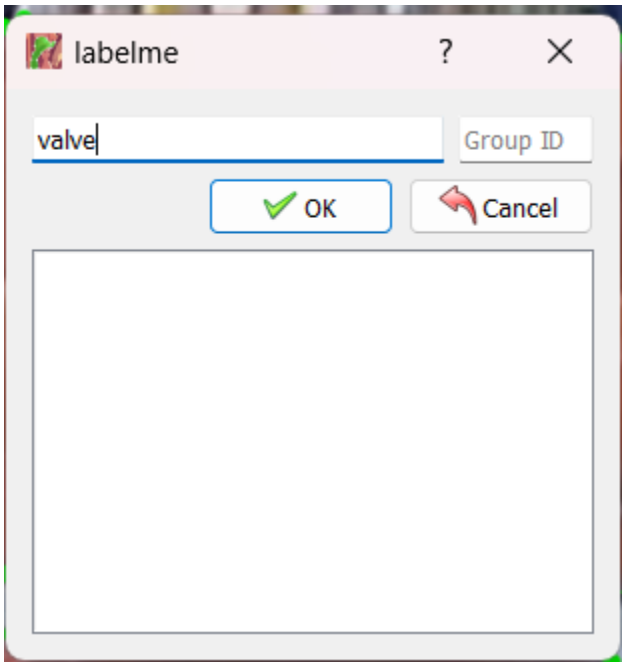


Figura 7. Ventana emergente para la asignación de tipo de etiqueta.

*Fuente: Labelme*

## 3.3 CONFIGURACIÓN DE ANACONDA Y TORCHSERVE

Después de completar el entrenamiento del modelo YOLO y una vez asignado el data set que contiene los datos con los que se va a reentrenar el modelo, se procede a cargar el modelo utilizando Anaconda y TorchServe para disponer de un servidor virtual de manera local en la ubicación específica donde va a trabajar este proceso se desarrolla de la siguiente manera:

### *3.3.1 Creación y Activación de un Entorno Virtual*

para garantizar la integridad y separación de las dependencias entre proyectos, se sugiere la creación de un entorno virtual dedicado utilizando Anaconda. Esto se logra mediante el

comando `conda create` como se ve en la figura 8, especificando la versión de Python deseada y el nombre del entorno virtual.

```
conda create -n torchserve_env python=3.8
```

Figura 8. Creción de ambiente virtual en torch serve.

*Fuente: CLI anaconda*

Una vez creado el entorno virtual, es fundamental activarlo antes de proceder con la instalación de las bibliotecas y herramientas necesarias. La activación se realiza a través del comando `conda activate` como se ve en la figura 9.

```
conda activate torchserve_env
```

Figura 9. Activación ambiente virtual de torch serve.

*Fuente: CLI anaconda*

### *3.3.2 Instalación de PyTorch y TorchServe*

PyTorch es una biblioteca popular para el desarrollo de modelos de IA en Python, y TorchServe es una biblioteca para disponer un servidor de modelos de PyTorch para producción, estas son las principales dependencias para que sea posible correr el modelo de manera local. Utilizando el gestor de paquetes conda, se instalan las versiones compatibles de PyTorch y TorchServe, asegurando la coherencia entre las bibliotecas. La instalación de estas bibliotecas se puede ejecutar simplemente escribiendo en la ventana de comandos que este usando para el proceso el comando pip install y el nombre de la biblioteca.

### *3.3.3 Carga del Modelo YOLO*

Una vez establecido el servidor local, se procede a descargar todas las dependencias necesarias para que el modelo corra, por ejemplo, torchvision, openCV, matplotlib, pandas, seaborn entre otras; esto puede ser descargado con un archivo de requerimientos previamente configurado para agilizar el proceso de descarga de dependencias. Luego se carga el modelo de YOLO con el que se trabajará.

### *3.3.4 Empaquetado del modelo de inteligencia artificial*

utilizando la herramienta `torch-model-archiver` como se ve en la figura 10, se procede a empaquetar el modelo YOLO en un formato compatible con TorchServe. Este paso implica la conversión del modelo de PyTorch a TorchScript o TorchScript Lite, junto con la definición del controlador (handler) y otros archivos necesarios para su correcto funcionamiento.

```
torch-model-archiver --model-name yolov5 --version 1.0 --serialized-file
yolov5.pt --handler yolov5_handler.py --extra-files yolov5.yaml
```

Figura 10. Empaquetado de modelo YOLOv5 en Torchserve

*Fuente: CLI anaconda*

Es importante modificar el "handler" para asegurarse de que funcione correctamente con el modelo entrenado, pues este se encarga de gestionar las solicitudes de inferencia y debe estar configurado adecuadamente para cargar y ejecutar el modelo en el entorno de producción local, este archivo es crucial para que el modelo pueda ser utilizado de manera

efectiva y hace posible realizar detecciones en tiempo real en esa ubicación particular.

### *3.3.5 Inicio de TorchServe*

El modelo será empaquetado luego de la ejecución del comando de la figura 10 y se procede a iniciar el servidor TorchServe, mediante el comando correspondiente como se ve en la figura 11. Se especifica la ubicación del almacenamiento de modelos (`model-store`) y se carga el modelo YOLO empaquetado para su disponibilidad inmediata.

```
torchserve --start --model-store model_store --models yolov5.mar
```

Figura 11. Inicialización de modelo en torchserve con fotmato .mar.

*Fuente: CLI anaconda*

### *3.3.6 Pruebas del sistema de monitoreo y seguridad laboral*

Para verificar la configuración y el funcionamiento adecuado de TorchServe con el sistema de monitoreo y seguridad laboral, se realizan pruebas enviando solicitudes HTTP a través de la API REST proporcionada. Se emplean herramientas como `curl` o bibliotecas de Python como `requests` para enviar solicitudes y recibir predicciones del modelo, validando así su desempeño en producción.

Una vez implementado el proceso descrito anteriormente, se realizan pruebas del sistema de monitoreo y seguridad laboral, donde se realizó la implementación del mismo y se hacen los ajustes necesarios según se desenvuelva el modelo de IA, para reconocer los elementos necesarios para su correcto funcionamiento y garantizar la seguridad de los trabajadores, comprobando que se prestará un servicio de manera correcta y efectiva.

# 4 VENTAJAS Y DESVENTAJAS DEL PROCESO ACTUAL EN EL SISTEMA DE MONITOREO Y SEGURIDAD LABORAL DE SLB

## 4.1 VENTAJAS

Contar con un sistema de sistema de monitoreo y seguridad laboral basado en IA de manera local, ofrece varias ventajas como las que se enumeran a continuación:

### *4.1.1 Control total sobre la Infraestructura*

Tener el sistema de manera local permite a las empresas el control total sobre sus servidores y la infraestructura en la que opera su inteligencia artificial. Esto proporciona una ventaja significativa en términos de personalización y ajustes específicos según las necesidades internas. Además, la seguridad física de los datos y equipos puede ser manejada directamente, asegurando que todas las medidas necesarias para proteger la infraestructura se implementen de manera óptima y específica.

### *4.1.2 Reducción de dependencia de Internet*

Operar localmente significa que el sistema puede seguir funcionando sin interrupciones incluso si hay problemas de conectividad a Internet, lo que asegura una operación continua y fiable. Este enfoque también reduce la latencia en la transmisión y procesamiento de datos, ya que todo el procesamiento se realiza en las instalaciones de la empresa, eliminando retrasos potenciales causados por la transmisión de datos a través de redes externas.

### *4.1.3 Personalización y optimización*

La gestión interna de los sistemas permite a las empresas optimizar sus configuraciones de hardware y software para satisfacer sus aplicaciones específicas. Esta flexibilidad puede no estar tan fácilmente disponible en un entorno de nube compartido. Además, la integración con sistemas y bases de datos locales ya existentes es más sencilla, lo que facilita la implementación de soluciones de inteligencia artificial sin necesidad de complejas adaptaciones.

### *4.1.4 Costos operativos predecibles*

Mantener los sistemas de manera local elimina la necesidad de pagar suscripciones mensuales o anuales a proveedores de servicios en la nube, lo que puede hacer que los costos sean más predecibles y controlables a largo plazo. Las empresas pueden maximizar el uso de la infraestructura existente sin tener que realizar nuevas inversiones significativas, aprovechando al máximo los recursos ya disponibles.

### *4.1.5 Protección de datos sensibles*

La gestión de datos de manera local permite a las empresas cumplir con las normativas y regulaciones locales sobre la protección de datos con mayor facilidad. Mantener el control directo sobre la privacidad y seguridad de los datos elimina la dependencia de terceros, asegurando que todas las medidas de seguridad y privacidad se apliquen estrictamente según los estándares internos y regulaciones externas.

### *4.1.6 Menor complejidad de migración de datos y tecnologías*

Mantener el sistema localmente evita la complejidad y los riesgos asociados con la migración de grandes volúmenes de datos a la nube. Esto asegura la continuidad del servicio sin riesgo de interrupciones significativas durante el proceso de migración. Las empresas pueden operar sin preocuparse por la pérdida de datos o la necesidad de adaptar nuevas infraestructuras, manteniendo la estabilidad y la integridad de su sistema.

## 4.2 DESVENTAJAS

Pese a las ventajas enumeradas anteriormente, tener un sistema de sistema de monitoreo y seguridad laboral basado en IA de manera local, conlleva ciertas limitaciones, por ejemplo:

### *4.2.1 Costos de mantenimiento y actualización*

Mantener un sistema de inteligencia artificial localmente implica costos continuos de mantenimiento y actualización de hardware y software. Las empresas deben invertir regularmente en la actualización de sus servidores y otros equipos para mantener el rendimiento óptimo y garantizar la compatibilidad con las últimas tecnologías. Estos costos pueden ser significativos y difíciles de prever, lo que impacta negativamente en el presupuesto de la organización.

### *4.2.2 Escalabilidad limitada para satisfacer los requerimientos futuros de la empresa*

Una infraestructura local limita la capacidad de escalar el sistema de manera rápida y eficiente. Cuando la empresa necesita aumentar su capacidad de procesamiento debido a un crecimiento repentino o la incorporación de nuevos proyectos, la infraestructura local puede no ser suficiente. Esto requiere inversiones adicionales en hardware, lo que no solo es costoso sino también lleva tiempo, retrasando la implementación de nuevas funcionalidades o proyectos.

### *4.2.3 Falta de flexibilidad para adaptarse a nuevas tecnologías y servicios*

La implementación local puede ser menos flexible en comparación con las soluciones basadas en la nube. Las empresas pueden enfrentar dificultades para adaptar su infraestructura a las nuevas necesidades tecnológicas o de mercado, ya que los cambios significativos en el hardware y software requieren tiempo y recursos. Además, la integración de nuevas tecnologías puede ser más compleja y lenta, afectando la capacidad de la empresa para innovar y mantenerse competitiva.

### *4.2.4 Recursos humanos y especialización*

Operar y mantener una infraestructura local requiere personal especializado en IT y en mantenimiento de hardware y software. La contratación y capacitación de este personal puede ser costosa y llevar tiempo. Además, las empresas deben asegurarse de que sus equipos sepan de las últimas tecnologías y prácticas de seguridad, lo que implica un esfuerzo continuo de formación y actualización.

### *4.2.5 Riesgos de Seguridad de los datos y la información*

Aunque tener control físico sobre los servidores puede parecer una ventaja, también puede exponer a la empresa a riesgos de seguridad si no se implementan medidas adecuadas. Los servidores locales pueden ser vulnerables a fallos de hardware, desastres naturales, robos o ataques físicos. Además, mantener la seguridad cibernética actualizada requiere un esfuerzo constante y recursos dedicados para proteger los datos sensibles y las operaciones.

### *4.2.6 Gestión de Datos Compleja*

La gestión de grandes volúmenes de datos en una infraestructura local puede ser complicada y menos eficiente. Los sistemas locales pueden enfrentar limitaciones en el almacenamiento y la velocidad de procesamiento, lo que puede afectar el rendimiento del sistema de inteligencia artificial. Además, la sincronización y el acceso a los datos desde diferentes ubicaciones pueden ser más difíciles de manejar, especialmente en empresas multinacionales con operaciones distribuidas.

### *4.2.7 Menor acceso a innovación y tecnología de punta*

Las soluciones en la nube ofrecen acceso rápido a las últimas innovaciones y tecnologías de punta, como la inteligencia artificial avanzada, el aprendizaje automático, y las herramientas de análisis de datos. Al operar localmente, las empresas pueden perderse estas oportunidades, ya que las actualizaciones y mejoras deben ser gestionadas internamente, lo que puede retrasar la adopción de nuevas tecnologías y limitar las capacidades del sistema.

# 5 ANÁLISIS DE TECNOLOGÍAS Y SERVICIOS EN LA NUBE PARA LA PROPUESTA DE ARQUITECTURA EN LA NUBE

En el contexto de la propuesta para la empresa SLB, es fundamental identificar las tecnologías y servicios disponibles en la nube que puedan facilitar la implementación del sistema de monitoreo y seguridad laboral basado en ML en SLB. Este capítulo se centra en la importancia de seleccionar adecuadamente estas tecnologías y servicios para garantizar un despliegue eficiente, seguro y escalable del del sistema de monitoreo y seguridad laboral. La correcta elección de herramientas en la nube no solo optimiza el rendimiento y la seguridad del sistema, sino que también reduce costos operativos y mejora la flexibilidad para adaptarse a nuevas necesidades empresariales.

La computación en la nube se refiere a la entrega de servicios informáticos, como servidores, almacenamiento, bases de datos, redes, software, análisis y más, a través de Internet ("la nube"). Este enfoque permite a las empresas acceder a recursos tecnológicos escalables y flexibles sin la necesidad de mantener una infraestructura física propia. Los beneficios generales incluyen la capacidad de escalar recursos según la demanda, flexibilidad en la implementación de nuevas soluciones y una significativa reducción de costos operativos al eliminar la necesidad de grandes inversiones en hardware y mantenimiento. Dentro de la computación en la nube, existen varios tipos de Servicios en la Nube, estos se clasifican principalmente en tres categorías:

### *5.1.1 Infrastructure as a service (IaaS)*

La infraestructura como servicio (IaaS) proporciona recursos informáticos virtualizados a través de Internet, como servidores, almacenamiento y redes, que los usuarios pueden aprovisionar y gestionar según sus necesidades. Este modelo es ideal para organizaciones

que requieren flexibilidad total sobre sus recursos de TI sin tener que invertir en infraestructura física. Google compute engine (GCE) para crear y gestionar máquinas virtuales de alto rendimiento que se emplearán en el entrenamiento del modelo de ML (YOLO). Esto permitirá flexibilidad en la configuración de los recursos computacionales necesarios para manejar grandes volúmenes de datos y realizar complejos cálculos de entrenamiento [37].

### *5.1.2 Platform as a service (PaaS)*

La plataforma como servicio (PaaS) ofrece un entorno completo que incluye hardware, software y herramientas para el desarrollo, despliegue y gestión de aplicaciones, permitiendo a los desarrolladores centrarse en la creación de aplicaciones sin preocuparse por la gestión de la infraestructura subyacente. En este contexto, se utilizará Google AI Platform Notebooks para desarrollar y experimentar con el modelo de ML, facilitando la colaboración y la iteración rápida sobre diferentes versiones del modelo. Además, Google Cloud AutoML permitirá simplificar el proceso de reentrenamiento del modelo, integrando datos específicos de la empresa y mejorando continuamente la precisión del modelo [37].

### *5.1.3 Software as a service (SaaS)*

El software como servicio (SaaS) permite acceder a aplicaciones software a través de Internet, eliminando la necesidad de instalaciones locales y reduciendo la carga de mantenimiento para los usuarios. Aunque el enfoque principal será el uso de modelos personalizados de ML, la Vision API de Google puede complementar las capacidades de reconocimiento de objetos, proporcionando una solución rápida y efectiva para analizar imágenes en tiempo real cuando sea necesario. Al combinar IaaS, PaaS y SaaS, la arquitectura en la nube será robusta, flexible y escalable, permitiendo una implementación

eficiente del modelo de ML en múltiples ubicaciones globales de SLB, optimizando así la seguridad laboral y la eficiencia operativa en la empresa [37].

## 5.2 HERRAMIENTAS DISPONIBLES EN LA NUBE PARA LA MEJORA DEL SISTEMA DE MONITOREO Y SEGURIDAD LABORAL EN SLB

### *5.2.1 Google cloud storage (GCS)*

Proporciona una solución de almacenamiento en la nube altamente escalable y duradera para datos de todo tipo, incluidos imágenes y grabaciones [5]. En el contexto la propuesta, GCS se utilizará para almacenar de manera segura y eficiente los grandes volúmenes de datos generados por las cámaras CCTV en las diferentes ubicaciones de SLB, gracias a su alta disponibilidad, GCS asegura que los datos estén accesibles en todo momento, mientras que su durabilidad garantiza la integridad y la preservación de los datos a largo plazo. Además, la escalabilidad de GCS permite manejar el crecimiento de datos sin necesidad de preocuparse por la capacidad de almacenamiento.

### *5.2.2 Google cloud Pub/Sub*

Es un sistema de mensajería en tiempo real diseñado para la ingestión y el procesamiento de datos. Este servicio será fundamental para gestionar el flujo continuo de datos que proviene de las cámaras CCTV, facilitando la transmisión de datos en tiempo real desde las cámaras hasta los servicios de procesamiento y almacenamiento en la nube. Su capacidad para manejar el procesamiento asíncrono permite que los datos sean analizados y utilizados para el reentrenamiento del modelo de ML sin demoras, mejorando la eficiencia y la capacidad de respuesta del sistema [7].

### *5.2.3 Google compute engine (GCE)*

Ofrece máquinas virtuales de alto rendimiento que se pueden personalizar según las necesidades específicas del sistema de monitoreo y seguridad laboral. En este caso, GCE se utilizará para el entrenamiento intensivo del modelo de ML YOLO. La escalabilidad de GCE permite ajustar los recursos computacionales en función de la carga de trabajo, lo que es crucial para manejar los complejos cálculos y el procesamiento de grandes data sets durante el entrenamiento del modelo. Además, la capacidad de personalización de GCE asegura que se puedan optimizar los recursos para obtener el mejor rendimiento y eficiencia en el entrenamiento del modelo, contribuyendo a mejorar la precisión y la velocidad de las inferencias.[1]

### *5.2.4 Google AI platform*

Es una plataforma integral diseñada para facilitar el desarrollo, entrenamiento y despliegue de modelos de ML. En el proyecto, esta plataforma se utilizará para entrenar el modelo YOLO con datos provenientes de los sistemas CCTV. La AI Platform ofrece integración con otras herramientas de Google Cloud, proporcionando un entorno cohesivo y eficiente para el desarrollo del modelo. Además, soporta frameworks populares como TensorFlow y PyTorch, lo que permite a los desarrolladores trabajar con tecnologías de vanguardia y optimizar el rendimiento del modelo [38].

### *5.2.5 Google kubernetes engine (GKE)*

es un servicio de orquestación de contenedores que permite desplegar y gestionar aplicaciones en contenedores de manera automatizada y escalable. En el contexto de la propuesta, GKE se utilizará para desplegar el modelo de ML entrenado en un entorno de

producción. GKE ofrece escalabilidad automática, lo que asegura que el sistema pueda manejar incrementos en la carga de trabajo sin problemas. También proporciona gestión automatizada y alta disponibilidad, garantizando que el modelo esté siempre accesible y funcione de manera eficiente, incluso en situaciones de alta demanda [6].

### *5.2.6 Google cloud functions y cloud run*

son servicios sin servidor que permiten ejecutar funciones y contenedores en una plataforma gestionada por Google. Estos servicios serán utilizados en la propuesta para implementar la lógica de procesamiento en tiempo real y la ejecución de tareas específicas relacionadas con el modelo de ML. Google Cloud Functions ofrece flexibilidad al permitir la ejecución de código en respuesta a eventos, mientras que Cloud Run permite ejecutar contenedores completos en una infraestructura sin servidor. Ambos servicios ofrecen un modelo de pago por uso, lo que ayuda a optimizar los costos operativos al pagar solo por los recursos consumidos [39].

### *5.2.7 Google cloud SQL y bigquery*

son servicios de almacenamiento y análisis de datos estructurados que proporcionan soluciones gestionadas para manejar grandes volúmenes de datos. Google Cloud SQL se utilizará para almacenar información estructurada sobre los datos capturados por el sistema de CCTV, incluyendo detalles de detección y metadatos [40]. BigQuery, por su parte, se empleará para realizar análisis rápidos y profundos de estos datos, permitiendo extraer insights valiosos y generar reportes detallados sobre el desempeño del modelo de ML. Estos servicios gestionados ofrecen alta disponibilidad y rendimiento, facilitando el manejo eficiente de datos y el análisis en tiempo real [41].

## 5.3 TECNOLOGÍAS DISPONIBLES PARA EL MEJORAMIENTO DEL MODELO DE IA

### *5.3.1 Tensorflow*

TensorFlow es un framework de código abierto desarrollado por Google para el desarrollo y entrenamiento de modelos de ML y deep learning. Es altamente flexible y escalable, permitiendo a los desarrolladores crear modelos desde simples hasta complejas arquitecturas de redes neuronales. En el contexto de este proyecto, TensorFlow será utilizado para el entrenamiento y reentrenamiento del modelo YOLO. Su capacidad para manejar grandes volúmenes de datos y su integración con otras herramientas de GCP lo hacen ideal para mejorar la precisión y eficiencia del modelo de detección de objetos [42].

### *5.3.2 PyTorch*

PyTorch es otro framework de código abierto muy popular para ML y deep learning, desarrollado por Facebook. Conocido por su facilidad de uso y su dinamismo, permite una mayor flexibilidad en la creación y experimentación con modelos de aprendizaje profundo. PyTorch proporciona una plataforma para experimentar con diferentes arquitecturas de modelos y técnicas de entrenamiento. Esto permitirá comparar resultados y optimizar el rendimiento del modelo YOLO, asegurando que se utilicen las mejores prácticas y enfoques para la detección de objetos [43].

### *5.3.3 Torchserve*

TorchServe es un sistema de servidor de modelos para PyTorch que facilita la implementación y el servicio de modelos de ML. Proporciona una infraestructura flexible para desplegar modelos entrenados en PyTorch, permitiendo la inferencia en tiempo real y la gestión de modelos de manera eficiente. TorchServe se utilizará para desplegar el

modelo YOLO una vez que haya sido entrenado. Esto permitirá a SLB ejecutar el modelo en producción, proporcionando inferencias en tiempo real para la detección de objetos en las grabaciones de CCTV. La capacidad de TorchServe para manejar múltiples modelos y versiones facilitará la actualización y mantenimiento del sistema de detección [43].

### *5.3.4 YOLO (You Only Look Once)*

YOLO es un modelo de detección de objetos conocido por su rapidez y precisión. A diferencia de otros métodos de detección, YOLO se basa en una sola red neuronal para predecir las clases y las posiciones de los objetos en una imagen. Hay varias versiones de YOLO, cada una mejorando en velocidad y precisión. YOLO, en particular, será el modelo central que se entrenará y desplegará para la detección de personas y objetos en sistema de monitoreo y seguridad laboral. Su implementación permitirá identificar de manera precisa y rápida si los trabajadores están usando el equipo de protección adecuado y siguiendo los protocolos de seguridad, mejorando así la seguridad laboral y la eficiencia operativa en SLB [17].

### *5.3.5 COCO data set*

COCO (Common Objects in Context) data set, por un lado, el data set COCO será usado como uno de los principales conjuntos de datos para este propósito, ya que este data set es conocido por su diversidad y riqueza, conteniendo más de 330,000 imágenes con 1.5 millones de instancias de objetos etiquetados en 80 categorías diferentes. Con el modelo de detección de objetos se puede aprender a identificar muchos objetos en contextos diversos, lo que aumenta su capacidad para detectar y clasificar objetos en nuevas imágenes, permitiendo a los modelos de ML aprender a identificar y localizarlos con gran precisión. En este proyecto, el COCO Data set se empleará como una de las principales fuentes de datos para entrenar y reentrenar el modelo YOLO. Su riqueza y diversidad en

anotaciones proporcionarán una base sólida para que el modelo aprenda a detectar objetos en las grabaciones de CCTV, mejorando la precisión y efectividad del sistema de monitoreo de seguridad en SLB [45].

### *5.3.6 Ultralytics*

Ultralytics es una empresa que desarrolla herramientas y frameworks para la implementación de modelos de Computer Vision, incluyendo la familia de modelos YOLO (You Only Look Once). Sus soluciones están diseñadas para facilitar el entrenamiento, reentrenamiento y despliegue de modelos de detección de objetos de alto rendimiento. Ultralytics proporcionará el framework y las herramientas necesarias para entrenar y desplegar el modelo YOLO. Esto incluye optimizaciones y mejoras específicas que hacen que el modelo sea más rápido y preciso en la detección de objetos en las grabaciones de CCTV, ayudando a SLB a implementar una solución de monitoreo de seguridad más efectiva [46].

### *5.3.7 CVAT - Roboflow:*

CVAT (Computer Vision Annotation Tool) es una herramienta de anotación de imágenes para entrenar modelos de Computer Vision. CVAT se utilizará para anotar manualmente las imágenes y videos capturados por las cámaras CCTV, creando un conjunto de datos etiquetados para el entrenamiento del modelo YOLO.

Roboflow es una plataforma que facilita la gestión de datases y la preparación de datos para entrenar modelos de ML. Roboflow asistirá en la organización, limpieza y augmentación del data set, asegurando que los datos estén en el mejor formato posible para optimizar el entrenamiento y rendimiento del modelo [47].

### *5.3.8 Computer vision*

Computer Vision es un campo de la inteligencia artificial que permite a las computadoras interpretar y procesar imágenes y videos del mundo real de manera similar a como lo hace el ojo humano. Utiliza técnicas avanzadas para reconocer patrones, detectar objetos y analizar escenas en imágenes y videos. En este proyecto, la tecnología de Computer Vision se aplicará para analizar las entradas de imagen que provee el sistema CCTV. El objetivo es detectar y reconocer personas y objetos, así como verificar el uso correcto de equipos de protección y manipulación de herramientas. La implementación de modelos avanzados como YOLO permitirá mejorar significativamente la precisión y eficiencia del sistema de monitoreo de seguridad en SLB [48].

### *5.3.9 Google AutoML*

Google AutoML es una suite de herramientas que permite a los desarrolladores crear modelos de ML personalizados sin necesidad de tener un profundo conocimiento técnico en el campo. Utiliza técnicas avanzadas de AutoML para optimizar la selección de modelos y el ajuste de hiperparámetros. Google AutoML se empleará para facilitar el reentrenamiento del modelo de detección de objetos con nuevos datos específicos de las ubicaciones de SLB. La capacidad de AutoML para automatizar el proceso de selección y optimización del modelo permitirá mejorar continuamente la precisión del sistema de detección sin requerir intervención manual intensiva [49].

### *5.3.10 Hyperparameter tuning*

Hyperparameter tuning es el proceso de ajustar los hiperparámetros de un modelo de ML para mejorar su rendimiento. Los hiperparámetros son configuraciones que no se aprenden del entrenamiento del modelo, pero tienen un impacto significativo en su eficacia y precisión. En este proyecto, el ajuste de hiperparámetros se realizará para optimizar el rendimiento del modelo YOLO. Utilizando técnicas automatizadas y plataformas como Google AI Platform, se probarán diferentes combinaciones de hiperparámetros para encontrar la configuración que ofrezca la mejor precisión y eficiencia en la detección de personas y objetos en las grabaciones de CCTV de SLB [50].

### *5.3.11 Albumentations*

Albumentations es una biblioteca de código abierto diseñada para la augmentación de imágenes. Image augmentation se refiere a las técnicas utilizadas para aumentar el tamaño y la diversidad del conjunto de datos mediante la aplicación de transformaciones como rotaciones, recortes, y ajustes de brillo, entre otras. Albumentations se empleará para aplicar técnicas de Image Augmentation al conjunto de datos utilizado para entrenar el modelo YOLO. Al diversificar el data set con imágenes aumentadas, se mejorará la robustez y generalización del modelo, permitiendo que el sistema de detección de objetos sea más preciso y fiable en diferentes condiciones y escenarios dentro de las grabaciones de CCTV [51].

## 5.4 TECNOLOGÍAS PARA EL PROCESAMIENTO, SEGURIDAD Y MONITOREO DE DATOS PARA EL SISTEMA DE MONITOREO Y SEGURIDAD LABORAL.

### *5.4.1 Apache Kafka*

Apache Kafka es una plataforma de streaming distribuida que se utiliza para construir pipelines de datos en tiempo real. Kafka permite la publicación y suscripción a flujos de registros, el almacenamiento de registros de manera tolerante a fallos, y el procesamiento de estos flujos en tiempo real. En este proyecto, Apache Kafka se utilizará para la transmisión de datos en tiempo real desde el sistema CCTV a los sistemas de procesamiento. Esto permitirá la ingestión continua de imágenes y videos capturados, facilitando su procesamiento y análisis sin retrasos significativos, mejorando así la eficacia del monitoreo de seguridad [52].

### *5.4.2 Apache Spark*

Apache Spark es un motor de análisis unificado de código abierto diseñado para el procesamiento de datos a gran escala. Ofrece una API para la programación en Java, Scala, Python y R, y optimiza el procesamiento de datos tanto en memoria como en disco. Apache Spark será utilizado para el procesamiento y análisis de los grandes volúmenes de datos capturados por el sistema CCTV. Su capacidad para manejar y procesar datos en paralelo permitirá un análisis rápido y eficiente de las grabaciones, ayudando a identificar patrones y mejorar la precisión del modelo de detección de objetos [53].

### *5.4.3 Hadoop HDFS (Hadoop Distributed File System)*

Es un sistema de archivos distribuido diseñado para almacenar grandes conjuntos de datos de manera fiable y para gestionar el almacenamiento en clústeres de servidores. HDFS se utilizará para almacenar y gestionar los grandes volúmenes de datos generados por las cámaras CCTV. Al proporcionar una solución de almacenamiento distribuido, HDFS permitirá manejar eficazmente la ingesta masiva de datos y asegurar su disponibilidad para el procesamiento y análisis continuo [54].

### *5.4.4 Vault*

Vault by HashiCorp es una herramienta para gestionar secretos y proteger datos sensibles. Proporciona una interfaz segura para almacenar y acceder a credenciales, claves de cifrado y otros secretos. Vault se utilizará para manejar y proteger claves de cifrado y otros datos sensibles dentro del proyecto. Esto garantizará que las credenciales y la información crítica se mantengan seguras, permitiendo un manejo seguro de los datos y protegiendo la integridad y la confidencialidad de la información procesada y almacenada en el sistema [55].

### *5.4.5 TLS (Transport Layer Security)*

Son protocolos de seguridad diseñados para cifrar las comunicaciones en la red y garantizar que los datos transmitidos entre los servidores y los clientes sean seguros y confidenciales. En este proyecto, TLS se utilizará para asegurar las comunicaciones entre los distintos componentes del sistema, como la transmisión de datos desde las cámaras CCTV a los servidores en la nube y entre los servicios de procesamiento de datos. Esto garantizará la protección contra intercepciones y accesos no autorizados durante la transmisión de datos sensibles [56].

### *5.4.6 Prometheus*

Prometheus es un sistema de monitoreo y alerta de código abierto que recopila y almacenas métricas en formato de series temporales, proporcionando herramientas para consultas y generación de alertas basadas en estas métricas. Prometheus se utilizará para monitorear el rendimiento y estado del sistema de ML y sus componentes. Esto incluye el seguimiento de la carga de trabajo, el uso de recursos, la latencia de las solicitudes y otros indicadores

clave de rendimiento, permitiendo una rápida identificación y resolución de problemas operativos [57].

### *5.4.7 ELK Stack (Elasticsearch, Logstash, Kibana)*

ELK Stack es un conjunto de herramientas de código abierto que incluye Elasticsearch para la búsqueda y análisis de datos, Logstash para el procesamiento de logs y Kibana para la visualización de datos en tiempo real. En el proyecto, ELK Stack se utilizará para gestionar logs y visualizar métricas del sistema. Elasticsearch facilitará la indexación y búsqueda eficiente de datos de logs, Logstash procesará y transformará los logs entrantes, y Kibana proporcionará un panel interactivo para visualizar y analizar las métricas y logs del sistema, mejorando así la observabilidad y el diagnóstico de problemas en tiempo real [58].

## 5.5 TECNOLOGÍAS PARA LA ORQUESTACIÓN Y DESPLIEGUE DEL SISTEMA DE MONITOREO Y SEGURIDAD LABORAL

### *5.5.1 Docker*

Docker es una plataforma que permite desarrollar, enviar y ejecutar aplicaciones dentro de contenedores, que son entornos ligeros y portátiles que incluyen todo lo necesario para ejecutar una aplicación, como código, bibliotecas y dependencias. En el proyecto, Docker se utilizará para crear contenedores que empaquetan el modelo de ML y todas sus dependencias, incluidas las bibliotecas de Python, frameworks de ML como TensorFlow o PyTorch, y cualquier otra herramienta o software necesario para el funcionamiento del modelo. Esto facilitará la distribución y despliegue del modelo en diferentes entornos de ejecución, garantizando la consistencia y portabilidad de la aplicación [59].

### *5.5.2 Apache Airflow*

Apache Airflow es un sistema de orquestación de flujos de trabajo que permite automatizar y gestionar tareas complejas de procesamiento de datos en un entorno distribuido. Airflow será utilizado en el proyecto para programar y monitorizar flujos de trabajo relacionados con el procesamiento de datos y el entrenamiento de modelos de ML. Con Airflow, se pueden definir flujos de trabajo complejos que incluyan tareas como la ingestión de datos desde fuentes externas, la limpieza y transformación de datos, el entrenamiento de modelos, la evaluación de modelos y el despliegue de modelos en producción. Airflow proporcionará una interfaz intuitiva para la gestión y supervisión de estos flujos de trabajo, lo que facilitará el desarrollo y mantenimiento del sistema de ML [8].

### *5.5.3 Jenkins*

Jenkins es una herramienta de automatización de código abierto que se utiliza para automatizar el proceso de integración continua (CI) y el despliegue continuo (CD). Permite la automatización de tareas repetitivas relacionadas con la construcción, prueba y despliegue de software. Jenkins se utilizará en el proyecto para automatizar el proceso de integración y despliegue del sistema de ML. Se pueden configurar pipelines en Jenkins que definan los pasos necesarios para construir, probar y desplegar el modelo de ML, incluyendo la integración con herramientas de control de versiones como Git y la ejecución de pruebas automatizadas. Esto facilitará la entrega continua del software y garantizará la calidad y estabilidad del sistema [9].

### *5.5.4 Terraform*

Terraform es una herramienta de infraestructura como código (IaC) que se utiliza para automatizar la creación, configuración y gestión de la infraestructura de forma declarativa.

Permite definir la infraestructura como código utilizando un lenguaje específico y gestionarla de manera eficiente. Terraform se utilizará en el proyecto para definir y gestionar la infraestructura necesaria para ejecutar el sistema de ML en la nube. Se pueden crear y gestionar recursos de infraestructura en servicios en la nube como GCP de manera automatizada y reproducible. Esto facilitará la implementación y gestión de la infraestructura necesaria para ejecutar el sistema de ML, garantizando su disponibilidad y escalabilidad [10].

# 6 PROPUESTA DE ESTRATEGIA PARA EL DEPLIEGUE DE ARQUITECTURA EN LA NUBE

Este capítulo se centrará en describir los pasos y estrategias para desplegar y optimizar el sistema de monitoreo y seguridad laboral en la plataforma de computación en la nube GCP, para lograr una solución escalable, flexible y capaz de manejar grandes volúmenes de datos en tiempo real.

## 6.1 FASE DE SEGURIDAD Y GESTIÓN DE ACCESO A LOS USUARIOS AL SISTEMA DE MONITOREO Y SEGURIDAD LABORAL

La seguridad y gestión de datos sensibles son componentes cruciales en cualquier arquitectura en la nube, especialmente cuando se manejan modelos de ML y grandes volúmenes de datos como los que maneja SLB. Al analizar la Figura 12, se evidencia la manera de como cada usuario realizará el proceso de ingreso a la arquitectura dispuesta en la nube, de este modo, cuando un usuario intenta ingresar al sistema, su conexión se establece inicialmente a través de TLS para asegurar que todos los datos transmitidos estén encriptados y protegidos contra intercepciones, una vez establecida la conexión segura, la solicitud del usuario pasa a través del firewall, el cual actúa como una barrera de seguridad que monitorea y controla el tráfico de red, permitiendo solo las conexiones autorizadas.

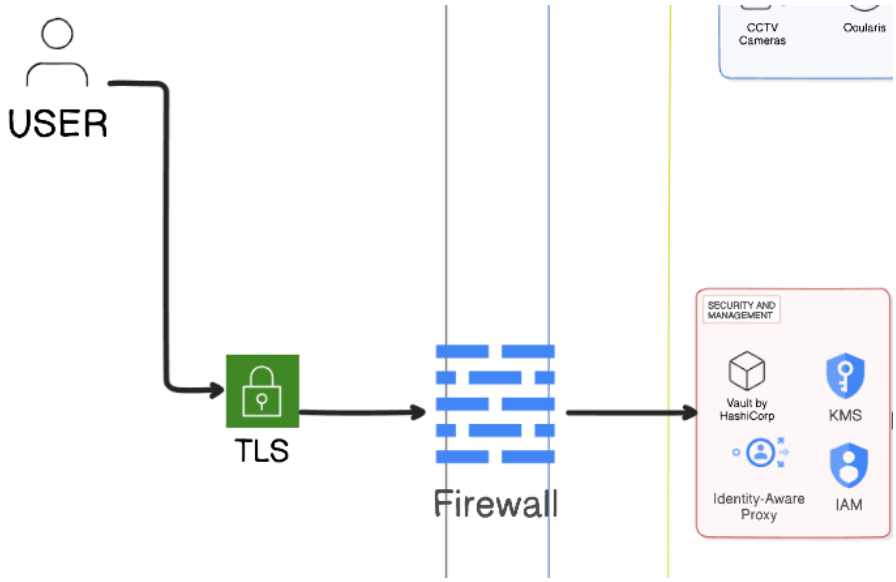


Figura 12. Área de acceso de usuario al sistema de monitoreo y seguridad laboral

*Fuente: Autor*

Después de pasar el firewall, la solicitud llega al bloque de Security and Management, donde se verifican las credenciales del usuario, es aquí donde Vault comienza a trabajar y se convierte en una herramienta fundamental en este aspecto, ya que ofrece una solución robusta para la gestión y control de acceso mediante tokens, contraseñas, certificados y claves de cifrado de manera centralizada y segura. Su capacidad para proporcionar acceso basado en políticas detalladas asegura que solo los servicios y usuarios autorizados puedan acceder a información crítica, reduciendo significativamente el riesgo de filtraciones y accesos no autorizados [54].

Configurar IAM en GCP, permite definir y gestionar quién tiene acceso a qué recursos, proporciona controles detallados de permisos a nivel de usuario y de servicio, asegurando que solo las personas autorizadas puedan acceder a los datos y servicios necesarios. Esto ayuda a minimizar los riesgos de acceso no autorizado y protege los datos sensibles.

El cifrado de las comunicaciones es otra pieza esencial de la seguridad en la nube, y aquí es donde TLS juega un papel vital, ya que cifran los datos transmitidos entre los diferentes componentes del sistema, como las comunicaciones entre las cámaras CCTV y los servidores de procesamiento, o entre los usuarios y la API de inferencia. Al cifrar los datos en tránsito, TLS garantiza que cualquier información intercambiada no pueda ser interceptada o manipulada por terceros malintencionados, protegiendo así la integridad y confidencialidad de los datos.

Para garantizar la protección de los datos en todas las etapas, es crucial implementar cifrado tanto en reposo como en tránsito, es aquí donde KMS entra en acción, haciendo posible gestionar claves de cifrado y realizar cifrado de datos almacenados (en reposo) en Google Cloud Storage buckets. Además, las políticas de seguridad de los buckets de almacenamiento aseguran que solo los usuarios y aplicaciones autorizados puedan acceder a los datos. El cifrado en tránsito se asegura de que los datos estén protegidos mientras se transfieren entre los servicios y usuarios, utilizando protocolos seguros como TLS, estas medidas combinadas aseguran un alto nivel de seguridad para los datos sensibles, protegiéndolos contra accesos no autorizados y posibles violaciones de datos.

El uso de IAP añade una capa adicional de seguridad al controlar y proteger las conexiones a la API de inferencia y otros servicios. IAP permite gestionar el acceso a las aplicaciones basadas en la identidad del usuario, proporcionando una autenticación centralizada y controles de acceso granulares [12]. esto es especialmente útil para proteger API ‘s y aplicaciones web, ya que asegura que solo los usuarios autenticados y autorizados puedan acceder a los recursos, independientemente de su ubicación. Con este último componente queda formado el bloque de seguridad y acceso a usuarios dentro de la arquitectura en la nube propuesta para SLB como se puede ver en la Figura 13.

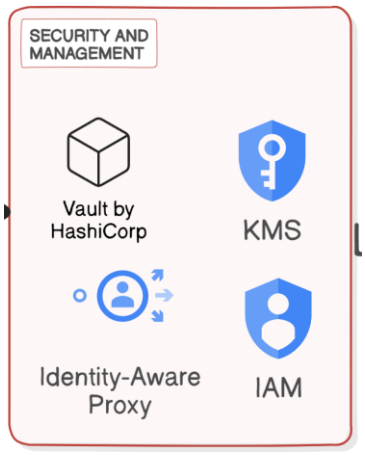


Figura 13. Bloque de seguridad y acceso de usuarios

*Fuente: Autor*

## 6.2 FASE DE PREPROCESAMIENTO DE DATOS

### *6.2.1 Captura de datos*

En esta fase, se implementará el sistema CCTV en cada ubicación de la empresa, estas zonas pueden abarcar áreas críticas de producción, almacenes, accesos, entre otras, con el fin de capturar imágenes y videos relevantes para el monitoreo y la seguridad de los trabajadores en la empresa. Junto al sistema CCTV, se utilizará el software Ocularis para acceder a las grabaciones de estas cámaras y transmitir los datos a la nube para su posterior procesamiento y análisis de los datos relevantes capturados. En la figura 14 se observa cómo quedará el bloque para la captura de datos.

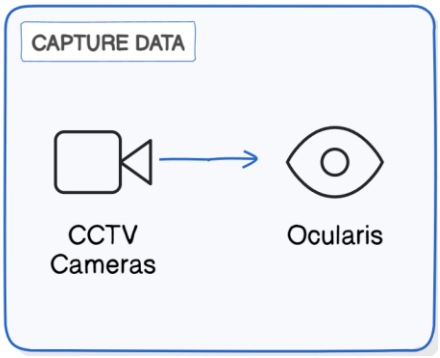


Figura 14. Bloque de sistema de captura de datos

*Fuente: Autor*

El software Ocularis se configurará para establecer conexión con las cámaras CCTV instaladas en las diferentes zonas de la ubicación en donde vaya a ser implementado el sistema de monitoreo y seguridad laboral, en la configuración de Ocularis, se pueden establecer reglas o acciones para que, cuando se capturen imágenes o videos en las cámaras CCTV, estos archivos se envíen automáticamente a GCS. Ocularis podría proporcionar una interfaz o configuración enfocada a especificar la dirección de almacenamiento de los eventos relevantes en GCS, así como las credenciales de autenticación necesarias para el acceso al servicio al que está intentando conectarse. Con esto, se podrá acceder a las grabaciones almacenadas en las cámaras y gestionar la transmisión de estos datos a la nube, la cual garantizará la disponibilidad de la información en tiempo real y permitirá tomar decisiones informadas basadas en la monitorización de las actividades de los trabajadores en las diferentes zonas de la empresa.

### *6.2.2 Almacenamiento de Datos*

El proceso de almacenamiento de datos para la implementación de soluciones de IA en la empresa se llevará a cabo mediante el uso de servicios clave de GCP, en primer lugar, se utiliza GCS, que es un servicio de almacenamiento en la nube altamente escalable y

duradero, para almacenar las imágenes y grabaciones provenientes del sistema CCTV. GCS proporcionara una infraestructura confiable para almacenar grandes volúmenes de datos de manera segura, organizada en buckets que pueden ser fácilmente gestionados y accedidos según las necesidades de la empresa.

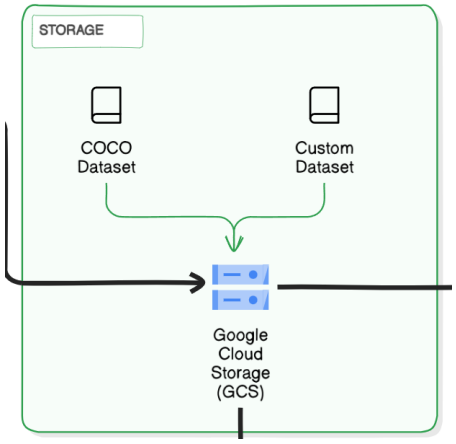


Figura 15. Bloque de captura y almacenamiento de datos

*Fuente: Autor*

El data set COCO Junto con el data set personalizado jugarán un papel fundamental en el proceso de implementación de esta propuesta de arquitectura en la nube, al utilizar el data set COCO, que contiene una amplia variedad de imágenes con anotaciones detalladas de objetos, junto con el data set personalizado que se enfoca en objetos y áreas específicas de la empresa, permiten que el modelo pueda mejorar su capacidad de reconocimiento al ser expuesto a una diversidad de escenarios y objetos. Complementariamente, al entrenar el modelo con data sets personalizados que reflejen los objetos y contextos específicos de la empresa, se garantiza una adaptación precisa a las necesidades del entorno laboral.

## 6.3 FASE DE ENTRENAMIENTO Y REENTRENAMIENTO DEL MODELO DE ML

### *6.3.1 Reentrenamiento con modelo de ML*

Teniendo en cuenta la idea anterior de los data sets que serán ubicados o creados (en el caso del data set personalizado) en el bloque de almacenamiento, el proceso de reentrenamiento del modelo se enfoca en mejorar la precisión y la capacidad de reconocimiento del modelo utilizando tanto la base de datos COCO como data sets personalizados de la empresa, el reentrenamiento con estos elementos es crucial para mejorar la precisión y la capacidad de generalización de los modelos de detección de objetos.

La combinación del COCO data set y data sets personalizados junto con las técnicas de aumentación de imágenes asegura que el modelo de detección de objetos se entrene de manera exhaustiva y robusta, ya que, al exponer al modelo a una mayor variedad de ejemplos durante el entrenamiento, se mejorará su capacidad para reconocer patrones y características distintivas de los objetos, incluso en condiciones adversas. Esto resulta en un modelo más preciso y confiable, capaz de operar eficazmente en escenarios del mundo real, como en las grabaciones de las cámaras CCTV utilizadas por SLB.

Por otro lado, GCP ofrece servicios especializados como Cloud AutoML y TensorFlow para facilitar el proceso de reentrenamiento con nuevos datos y etiquetas específicas de la empresa, es aquí donde entran herramientas como Cloud AutoML y TensorFlow, las cuales simplifican el proceso de reentrenamiento de modelos de IA al proporcionar herramientas avanzadas de ML y capacidades de automatización en la creación de modelos de ML, lo que permitirá que el equipo de trabajo de la empresa pueda realizar el reentrenamiento del modelo de manera eficiente y efectiva, aprovechando la infraestructura escalable en la nube y las herramientas especializadas que ofrece GCP.

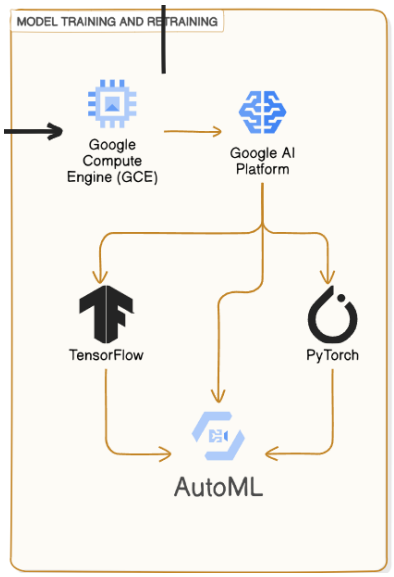


Figura 16. Bloque de entrenamiento y reentrenamiento modelo de ML

*Fuente: Autor*

### *6.3.2 Entrenamiento de modelos de ML*

El entrenamiento de modelos en esta arquitectura se inicia con GCE, que proporcionará máquinas virtuales de alto rendimiento necesarias para el entrenamiento intensivo de los modelos de IA que desee implementar SLB para esta y futuras soluciones dentro de la empresa. Estas máquinas virtuales ofrecen la capacidad de procesamiento y la flexibilidad para ejecutar tareas complejas de ML, lo cual posibilitará llevar a cabo el entrenamiento del modelo YOLO en la plataforma GCP, en donde entrarán los datos para el entrenamiento y reentrenamiento del modelo de ML tal como se observa en la figura 16, permitiendo el manejo eficiente de grandes data sets y la ejecución rápida de algoritmos de entrenamiento. Además, GCE asegura que los recursos computacionales se escalen según sea necesario, optimizando el tiempo y los costos asociados con el entrenamiento de modelos de IA.

Google AI platform es un complemento esencial para esta infraestructura ya que ofrece un entorno completo para el desarrollo, entrenamiento y despliegue de modelos de ML, con la ayuda de herramientas integradas para la gestión del ciclo de vida de los modelos facilita tareas como el preprocesamiento de datos, el ajuste de hiperparámetros y la validación de modelos. Además, Google AI Platform permite la colaboración entre equipos y la integración continua de mejoras, asegurando que los modelos estén siempre actualizados y listos para su despliegue en producción.

Los frameworks de ML, como TensorFlow y PyTorch, son fundamentales en este proceso de entrenamiento de modelos de ML, ya que proporcionan las bibliotecas y herramientas necesarias para el desarrollo y entrenamiento de los modelos. Por un lado, TensorFlow, con su amplia adopción y robustez, es ideal para la creación de modelos de ML complejos y la optimización de estos. Por otro lado, PyTorch, conocido por su flexibilidad y facilidad de uso, permite una experimentación rápida y eficiente con diferentes arquitecturas de modelos y técnicas de entrenamiento. Juntos, estos frameworks aseguran que los modelos de detección de objetos, como YOLO, se desarrollen y entrenen con la máxima precisión y eficiencia para solucionar problemas como el que se abarca en la presente monografía.

## 6.4 FASE DE DESPLIEGUE E IMPLEMENTACIÓN DEL MODELO DE ML

Para realizar el despliegue de modelos de ML de manera eficiente, habrá una combinación de varias herramientas que trabajarán en conjunto para lograr mejorar y optimizar el proceso a comparación de cómo se lleva actualmente, para esto es necesario contar con GKE, Docker y TorchServe. En primer lugar, GKE realizará la orquestación de contenedores, permitiendo gestionar y escalar automáticamente las aplicaciones y los modelos de ML en contenedores según la demanda, emplear contenedores es crucial para el correcto despliegue de uno o varios modelos de ML, de modo que al implementar esta

herramienta en los procesos de despliegue ayudará a adoptar buenas prácticas y minimizará errores. GKE asegura que los modelos de ML desplegados puedan manejar grandes volúmenes de solicitudes y ajustarse dinámicamente a las cargas de trabajo cambiantes, proporcionando una infraestructura robusta y escalable[3].

Docker jugará un papel crucial en el despliegue de aplicaciones, gracias a su capacidad de encapsular el software en contenedores, junto con todas sus dependencias, los modelos de ML y sus entornos de ejecución se empaquetarán en contenedores portátiles y consistentes, lo que facilita el despliegue de estos contenedores en diferentes entornos sin conflictos de dependencias [59]. Esto asegurará que el modelo entrenado se ejecute de manera idéntica en el entorno de desarrollo, prueba y producción, reduciendo significativamente los problemas de integración, minimizando errores y facilitando la colaboración entre equipos para el despliegue de este.

TorchServe será la herramienta especializada para desplegar modelos de ML entrenados con PyTorch en los entornos de desarrollo propuestos por la empresa, gracias a sus funcionalidades como servir modelos a través de API's RESTful, gestionar múltiples versiones de modelos y ayudar a escalar automáticamente en respuesta a la demanda de las solicitudes a los modelos de ML, se obtiene una solución potente y flexible para desplegar y gestionar modelos de ML en cada entorno de desarrollo, asegurando una alta disponibilidad y rendimiento eficiente, permitiendo a SLB desplegar rápidamente modelos de detección de objetos entrenados y adaptarlos conforme evolucionan las necesidades del negocio [43].

Apache Airflow será usado para la orquestación de flujos de trabajo, lo cual permite definir, programar y monitorizar tareas complejas de procesamiento de datos y entrenamiento de modelos, haciendo posible gestionar fácilmente las dependencias entre tareas, programar ejecuciones periódicas y manejar nuevos despliegues de los servicios en caso de fallos,

asegurando que todos los componentes del pipeline de datos y modelos de ML funcionen de manera coordinada y eficiente.

Jenkins es una herramienta clave para la integración y despliegue continuo (CI/CD), automatizando la construcción, prueba y despliegue de aplicaciones. En el contexto del ML, Jenkins facilita la integración continua de cambios en el código del modelo y su despliegue en entornos de prueba y producción, permitiendo a los equipos de desarrollo iterar rápidamente, validar nuevas versiones del modelo y desplegarlas con confianza, mejorando la agilidad y reduciendo el tiempo de entrega de nuevas funcionalidades y mejoras en los modelos de ML [60].

Terraform es una herramienta de Infraestructura como Código (IaC) que permite definir, provisionar y gestionar la infraestructura de manera programática [37]. Esta herramienta hará posible crear y gestionar de forma automatizada los recursos de la nube necesarios para el despliegue de modelos, como máquinas virtuales, redes, almacenamiento y servicios de contenedores. Esto asegura que la infraestructura sea reproducible, escalable y consistente, facilitando la gestión de entornos complejos y la implementación de cambios en la infraestructura de manera segura y eficiente. En conjunto, Apache Airflow, Jenkins y Terraform proporcionan una solución robusta y escalable para la automatización y orquestación de todo el ciclo de vida del desarrollo y despliegue de modelos de ML en la nube.

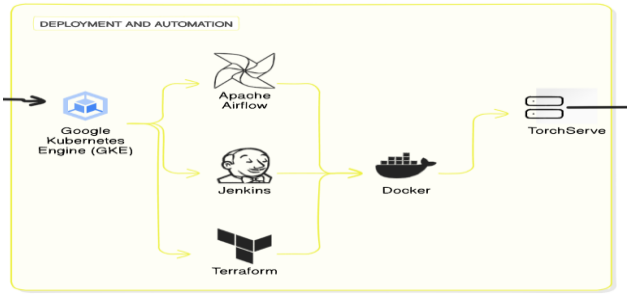


Figura 17. Bloque de despliegue y automatización de sistema de monitoreo y seguridad laboral

*Fuente: Autor*

## 6.5 FASE DE IMPLEMENTACIÓN DE PIPELINE

El pipeline del sistema funcionará mediante una serie de etapas automatizadas que aseguran el flujo continuo y eficiente de datos desde la captura hasta la inferencia. Comenzará con la captura de imágenes y videos a través de cámaras CCTV, cuyos datos se transmitirán en tiempo real para el procesamiento y análisis de datos con Apache Kafka, permitiendo la ingesta continua de grandes volúmenes de datos desde las cámaras CCTV hacia el modelo de ML. Kafka actuará como un sistema de mensajería distribuido, asegurando que los datos de video e imágenes capturados se transmitan de manera eficiente y en tiempo real a los sistemas de procesamiento. Este flujo constante de datos es crucial para aplicaciones que requieren análisis en tiempo real, como la detección de objetos y eventos de seguridad laboral.

Una vez que los datos se transmiten a través de Kafka, Apache Spark entra en acción para el procesamiento y análisis de los datos entrantes. Spark será el motor de análisis que manejará los grandes volúmenes de datos a gran velocidad y eficiencia en el procesamiento de los mismos, gracias a su capacidad de procesamiento en memoria, Spark permite realizar tareas de análisis complejas en el procesamiento de imágenes y la ejecución de algoritmos de ML, de manera mucho más rápida que las soluciones tradicionales basadas en disco [61]. Esto es esencial para la identificación rápida y precisa de personas, cascos, guantes y herramientas identificadas en las grabaciones del sistema CCTV.

Para el almacenamiento a largo plazo y la gestión de los grandes volúmenes de datos generados, se utiliza Hadoop HDFS, que actúa como un sistema de archivos distribuido, gracias a la capacidad de almacenar grandes cantidades de datos de manera redundante y distribuida, garantizando la disponibilidad y fiabilidad de los datos incluso en caso de fallos

de hardware [53]. Esta capacidad es fundamental para mantener un archivo completo de todas las grabaciones de CCTV y los resultados del análisis, proporcionando una base sólida para el reentrenamiento continuo de los modelos de ML y para cumplir con los requisitos de auditoría y estándares fijados por SLB.

Además, se implementa Cloud Pub/Sub, un servicio de mensajería en tiempo real de GCP, para crear un flujo de datos continuo que procesa las imágenes y videos capturados por las cámaras CCTV. Cloud Pub/Sub facilita la transmisión eficiente de mensajes entre las cámaras y los sistemas de procesamiento y análisis de datos en la nube. Mediante la suscripción a temas específicos, las aplicaciones pueden recibir y procesar los mensajes entrantes en tiempo real, lo que permite una rápida respuesta a los eventos detectados por las cámaras

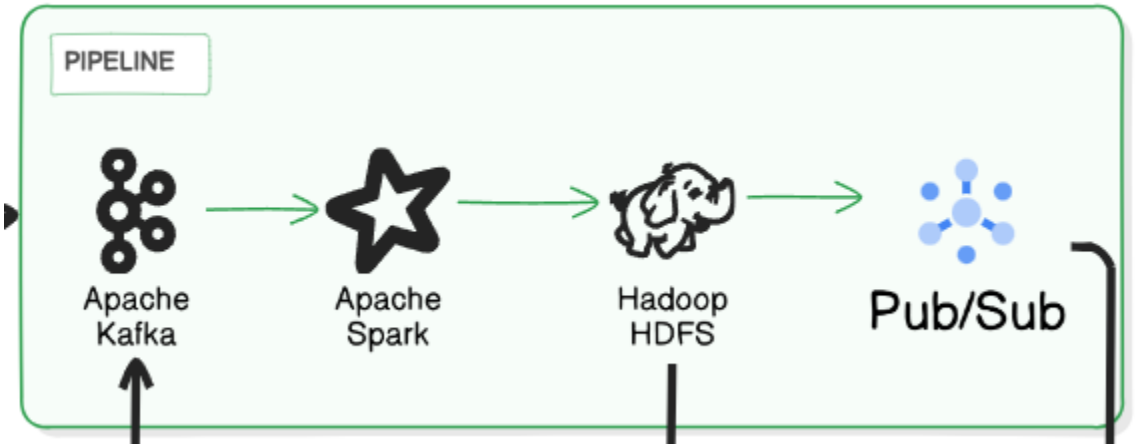


Figura 18. Bloque de pipeline

*Fuente: Autor*

## 6.6 FASE DE MONITOREO Y VISUALIZACIÓN DEL SISTEMA DE MONITOREO Y SEGURIDAD LABORAL

El monitoreo y la visualización son componentes críticos para garantizar la operatividad y eficiencia de una arquitectura en la nube. En primer lugar, Prometheus servirá como una herramienta de monitoreo y alerta de cambios, puesto que recopila y almacenas métricas de diversos servicios y aplicaciones, permitiendo a los administradores configurar alertas para ser notificados de manera proactiva sobre cualquier problema de rendimiento o fallos en el sistema. Esto facilita una rápida respuesta a incidentes y ayuda a mantener la estabilidad del sistema.

El ELK Stack, será usado para la búsqueda, análisis y visualización de datos en tiempo real. Por un lado, Elasticsearch sirve como motor de búsqueda y análisis distribuido, que permite indexar y buscar grandes volúmenes de datos de manera rápida y eficiente [57]. Logstash se encargará de la recolección, procesamiento y almacenamiento de logs, mientras que Kibana hará posible la visualización y creación de dashboards interactivos para monitorear el estado del sistema por parte de los administradores. El uso de ELK Stack en la arquitectura permite a los administradores tener una visión detallada y comprensible de los datos operacionales, facilitando la identificación de tendencias, problemas y oportunidades de optimización.

Google Cloud Functions y Cloud Run complementan estas herramientas al ofrecer capacidades de ejecución sin servidor. Google Cloud Functions permite ejecutar funciones en respuesta a eventos sin necesidad de gestionar servidores, lo que es ideal para tareas de procesamiento en tiempo real y para manejar solicitudes de inferencia de manera eficiente. Cloud Run permite desplegar contenedores de manera totalmente administrada y escalable, lo que facilita la ejecución de aplicaciones basadas en contenedores con una administración mínima. Estas herramientas permiten a los desarrolladores centrarse en el código y la lógica de la aplicación sin preocuparse por la infraestructura subyacente, mejorando la eficiencia y la agilidad en el desarrollo y despliegue de servicios.


Figura 19. Bloque de monitoreo y visualización

*Fuente: Autor*

Por último, la API REST se conectará al conjunto de monitoreo y visualización de la arquitectura a través de integraciones con herramientas como Prometheus y el stack ELK y demás, esta conexión permitirá que todos los análisis y resultados generados por los modelos de machine learning sean visibles en tiempo real en una interfaz gráfica intuitiva. A medida que el modelo procesa las imágenes y videos capturados, cualquier detección de objetos, alertas de seguridad, o anomalías serán enviadas a la API REST, estos datos se transmitirán a la plataforma de monitoreo, donde se almacenarán, analizarán y presentarán en dashboards interactivos. Los usuarios podrán ver resúmenes detallados, gráficos y alertas en tiempo real, permitiéndoles responder rápidamente a cualquier incidente detectado por el sistema.


Figura 20. API Rest

*Fuente: Autor*

De este modo, la arquitectura aquí propuesta, proporciona una base sólida para implementar un sistema de reconocimiento de personas y objetos utilizando tecnologías avanzadas en GCP.

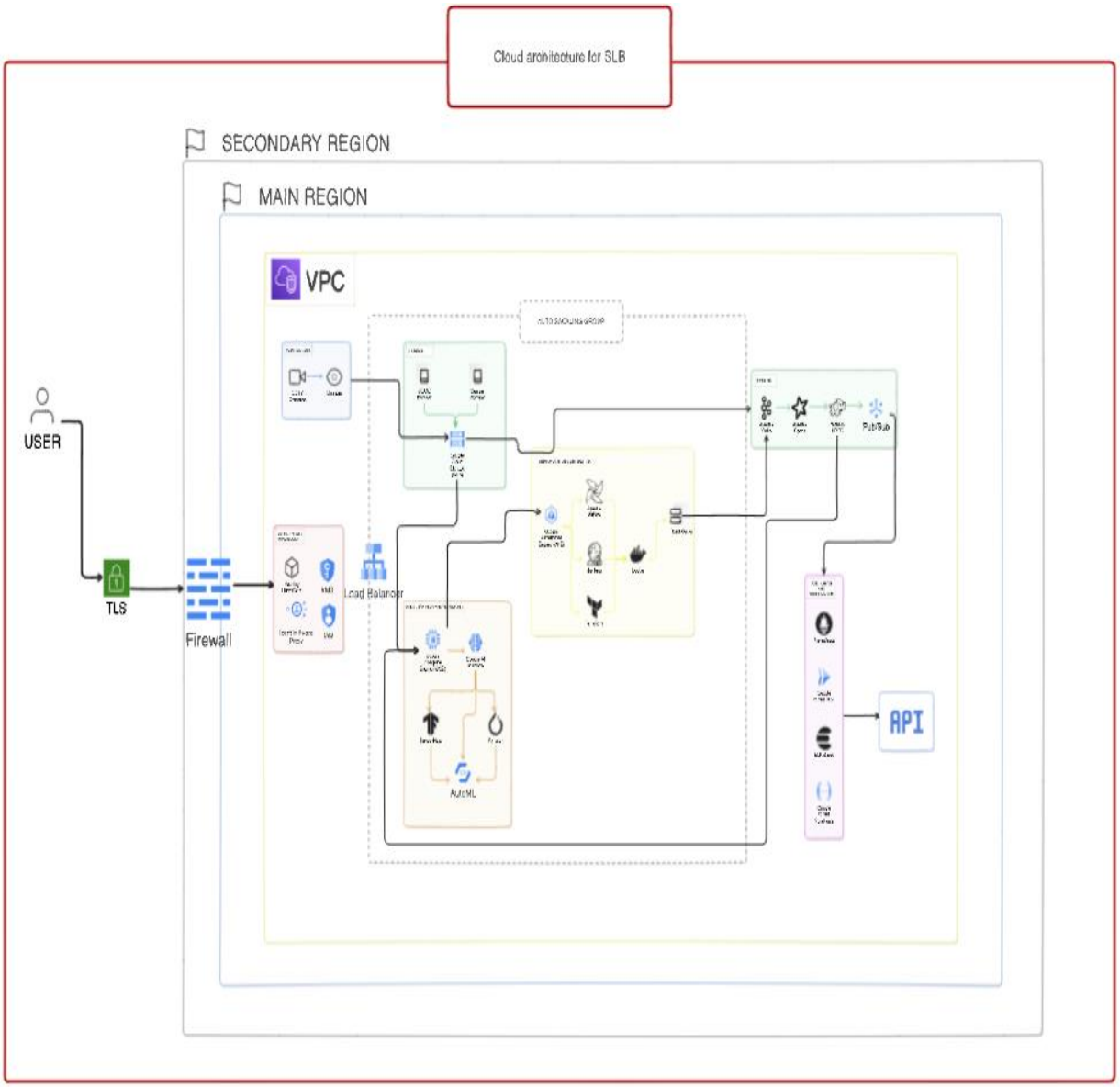


Figura 21. Propuesta de arquitectura en la nube para sistema de monitoreo y seguridad laboral en SLB

*Fuente: Autor*

# 7 BENEFICIOS DE UNA ARQUITECTURA EN LA NUBE PARA LA EMPRESA SLB

En la era digital actual, la adopción de una arquitectura en la nube se ha vuelto esencial para las empresas que buscan mantenerse competitivas y eficientes. La computación en la nube ofrece beneficios que van más allá de la capacidad de almacenamiento, incluyendo mejoras significativas en la eficiencia operativa, la escalabilidad y la seguridad. Para una empresa líder en el sector energético como SLB, aprovechar estas tecnologías no solo es una oportunidad para optimizar sus operaciones sino también para innovar en la forma en que gestiona sus recursos y procesos.

El objetivo de este capítulo es analizar los beneficios específicos que la adopción de una arquitectura en la nube puede ofrecer a SLB. Se examinarán aspectos clave como la eficiencia operativa, la escalabilidad y la seguridad, destacando cómo estas mejoras pueden impactar positivamente en las operaciones diarias y en la estrategia a largo plazo de la empresa.

## 7.1 EFICIENCIA OPERATIVA

### *7.1.1 Reducción de Costos*

El despliegue de una arquitectura en la nube puede reducir costos en comparación con la infraestructura local tradicional. En lugar de invertir en hardware costoso y su mantenimiento, SLB puede aprovechar los servicios en la nube que ofrecen un modelo de pago por uso. Esto no solo elimina los costos de capital iniciales, sino que también reduce los gastos operativos relacionados con la actualización y mantenimiento del hardware. Además, la nube permite una mejor previsión de costos, evitando gastos inesperados y mejorando la planificación financiera.

### *7.1.2 Optimización de Recursos*

La optimización de recursos es otro beneficio clave de la adopción de una arquitectura en la nube. Los servicios en la nube permiten el uso eficiente de recursos computacionales, ajustando automáticamente la capacidad según la demanda. Esto significa que SLB puede escalar recursos hacia arriba o hacia abajo en función de las necesidades operativas, evitando el subuso o sobreuso de recursos. La nube también facilita la automatización de procesos y tareas rutinarias, liberando tiempo y recursos para que los empleados se concentren en actividades de mayor valor.

### *7.1.3 Flexibilidad y Adaptabilidad*

La capacidad de ajustar recursos según la demanda es una ventaja crucial de la nube. SLB puede responder rápidamente a cambios en el mercado o en la demanda sin la necesidad de realizar inversiones significativas en infraestructura. Por ejemplo, durante los picos de actividad, la empresa puede aumentar temporalmente su capacidad computacional para manejar la carga adicional y luego reducirla una vez que la demanda disminuye. Esta flexibilidad permite a SLB mantener una operación ágil y adaptativa, mejorando su capacidad para competir en un entorno de negocios en constante cambio.

### *7.1.4 Automatización de Procesos*

La automatización de procesos es un beneficio significativo de la adopción de servicios en la nube. Los servicios en la nube permiten la automatización de tareas repetitivas y administrativas, reduciendo la carga de trabajo manual y minimizando la posibilidad de errores humanos. Por ejemplo, tareas como la gestión de datos, la administración de recursos y la implementación de actualizaciones pueden ser automatizadas, lo que no solo

mejora la precisión, sino que también aumenta la productividad. Para SLB, esto significa que los empleados pueden centrarse en actividades más estratégicas y de mayor valor, mientras que los sistemas en la nube gestionan las operaciones rutinarias de manera eficiente.

## 7.2 ESCALABILIDAD

### *7.2.1 Escalabilidad Vertical y Horizontal*

La escalabilidad es una de las características más importantes de una arquitectura en la nube. La escalabilidad vertical implica aumentar la capacidad de los recursos existentes, como agregar más memoria RAM o CPU's a una máquina virtual. La escalabilidad horizontal, por otro lado, consiste en agregar más máquinas o nodos al sistema. SLB puede beneficiarse de la escalabilidad vertical para mejorar el rendimiento de aplicaciones críticas sin interrumpir el servicio, mientras que la escalabilidad horizontal es ideal para manejar aumentos significativos en la carga de trabajo, como el procesamiento de grandes volúmenes de datos provenientes de las cámaras CCTV.

### *7.2.2 Gestión de Picos de Trabajo*

La capacidad de manejar cargas de trabajo fluctuantes es crucial para una empresa global como SLB. La arquitectura en la nube permite ajustar dinámicamente los recursos en respuesta a picos de trabajo, asegurando un rendimiento óptimo sin necesidad de mantener una infraestructura sobredimensionada permanentemente. Por ejemplo, durante auditorías de seguridad o eventos de alto riesgo, SLB puede escalar rápidamente sus recursos para procesar y analizar grandes cantidades de datos en tiempo real, garantizando una operación continua y segura.

### *7.2.3 Expansión Global*

Una de las ventajas más destacadas de la nube es la facilidad para desplegar servicios en múltiples regiones geográficas. Esto permite a SLB expandir sus operaciones de manera eficiente y rápida, aprovechando la infraestructura global de los proveedores de nube. Con servicios distribuidos en varias regiones, SLB puede mejorar la redundancia, reducir la latencia y proporcionar un mejor rendimiento a nivel mundial. Esta capacidad es beneficiosa para coordinar operaciones en diferentes lugares y garantizar que las filiales trabajen de forma sincronizada y eficiente.

## 7.3 SEGURIDAD

### *7.3.1 Protocolos de Seguridad Avanzados*

Los proveedores de servicios en la nube implementan medidas de seguridad avanzadas para proteger los datos y aplicaciones. Estos incluyen cifrado de datos en tránsito y en reposo, firewalls de última generación, y sistemas de detección y prevención de intrusiones. Para SLB, esto significa una protección robusta de los datos sensibles, asegurando que la información crítica esté siempre protegida contra accesos no autorizados y ciberataques. Por ejemplo, el cifrado robusto garantiza que los datos críticos de los empleados y las operaciones de la empresa permanezcan seguros incluso en caso de una brecha de seguridad.

### *7.3.2 Gestión de Accesos y Permisos*

La implementación de controles de acceso detallados permite a SLB gestionar de manera eficiente quién tiene acceso a qué datos y recursos. Los servicios en la nube permiten la implementación de controles de acceso detallados y la gestión centralizada de permisos,

incluyendo los sistemas de gestión de identidad y acceso (IAM) en la nube, facilitando la administración centralizada de permisos y asegurando que solo el personal autorizado pueda acceder a información confidencial.

SLB puede utilizar estas capacidades para definir roles y permisos específicos para diferentes empleados y equipos, asegurando que cada usuario tenga acceso únicamente a los recursos necesarios para su trabajo, reduciendo así el riesgo de errores y brechas de seguridad. Esto no solo aumenta la seguridad, sino que también simplifica el cumplimiento de las políticas internas y regulaciones externas.

### *7.3.3 Resiliencia y Recuperación ante Desastres*

La capacidad de recuperación ante desastres es crucial para cualquier estrategia de continuidad del negocio, en este caso, los proveedores de servicios en la nube ofrecen soluciones avanzadas de recuperación ante desastres que permiten a SLB restaurar rápidamente los sistemas y datos críticos en caso de incidente. Estas soluciones incluyen la replicación de datos en múltiples regiones y la creación de copias de seguridad automatizadas, lo que asegura que SLB pueda continuar sus operaciones con mínima interrupción.

### *7.3.4 Protección de Datos*

La protección de datos es fundamental para mantener la confianza y cumplir con las normativas de privacidad. Los proveedores de servicios en la nube implementan medidas de seguridad avanzadas, como el cifrado de extremo a extremo y auditorías de seguridad regulares para proteger los datos sensibles. SLB puede utilizar estas características para

garantizar que la información confidencial de la empresa y sus empleados esté protegida contra posibles amenazas.

### *7.3.5 Cumplimiento y Normativas*

Cumplir con las regulaciones y normativas de la industria es esencial para operar de manera legal y ética. Los servicios en la nube ayudan a SLB a cumplir con las normativas específicas del sector, como las leyes de protección de datos y las normas de seguridad. Los proveedores de nube suelen estar certificados por estándares reconocidos internacionalmente, lo que facilita a SLB demostrar cumplimiento y mantener la confianza de sus clientes y socios.

## 7.4 IMPACTO EN MODELOS DE ML

### *7.4.1 Mejora en la Precisión*

La adopción de una arquitectura en la nube permite a SLB usar data sets más grandes y variados, lo que contribuye a mejorar la precisión de los modelos de ML. Con acceso a muchos datos, los modelos aprenden patrones más complejos y detallados, reduciendo los errores y aumentando la exactitud de las predicciones. Por ejemplo, al entrenar modelos de detección de objetos, SLB puede utilizar datos globales en tiempo real para mejorar la precisión en la identificación de riesgos laborales.

### *7.4.2 Generalización de Resultados*

La capacidad de entrenar modelos más robustos y generalizables es otra ventaja importante de utilizar la nube. Al aprovechar recursos computacionales avanzados y diversos data sets, SLB puede desarrollar modelos que no solo funcionen bien en entornos controlados, sino que también mantengan su desempeño en situaciones reales y variadas. Esto es crucial para la implementación práctica de los modelos, asegurando que las soluciones de ML sean efectivas en diferentes contextos operacionales y geográficos.

### *7.4.3 Reducción de Tiempos de Entrenamiento*

El uso de recursos computacionales avanzados en la nube acelera significativamente el proceso de entrenamiento de los modelos de ML. Al aprovechar la potencia de cómputo escalable y las capacidades de procesamiento paralelo, SLB puede reducir drásticamente los tiempos de entrenamiento y reentrenamiento. Esto no solo disminuye los costos asociados al tiempo de computación, sino que también permite una iteración más rápida y eficiente en el desarrollo de modelos. Por ejemplo, los tiempos de entrenamiento pueden reducirse de semanas a días, facilitando una actualización continua y rápida de los modelos en respuesta a nuevos datos.

## 7.5 CASOS DE USO ESPECÍFICOS EN SLB

### *7.5.1 Monitoreo de Seguridad Laboral*

La implementación de una arquitectura en la nube mejora significativamente el monitoreo de seguridad laboral en SLB. Los modelos de ML pueden detectar de manera eficiente incumplimientos de seguridad, como el uso incorrecto de equipos de protección personal o la presencia de personas en áreas restringidas. Estos sistemas permiten una respuesta

rápida a incidentes, minimizando riesgos y mejorando la seguridad general en el lugar de trabajo. La capacidad de monitorear y analizar datos en tiempo real asegura que las condiciones laborales cumplan con las normativas y estándares de seguridad.

### *7.5.2 Gestión de Datos y Análisis Predictivo*

La gestión eficiente de grandes volúmenes de datos es fundamental para el análisis predictivo. Con la nube, SLB puede almacenar y procesar datos a gran escala, permitiendo el desarrollo de modelos predictivos avanzados. Estos modelos pueden predecir tendencias y potenciales problemas antes de que ocurran, facilitando una toma de decisiones proactiva y mejor informada. La capacidad de analizar datos en tiempo real mejora la precisión de las predicciones y optimiza la planificación operativa.

## 7.6 IMPACTO EN SLB

### *7.6.1 Aumento de la Competitividad*

La adopción de una arquitectura en la nube puede otorgar a SLB una ventaja competitiva significativa. La capacidad de escalar rápidamente, acceder a recursos avanzados y mejorar la precisión de los modelos de ML permite a SLB sobresalir frente a sus competidores. Por ejemplo, en el sector de la tecnología y los servicios, SLB puede ofrecer soluciones más innovadoras y eficientes, atrayendo a más clientes y fortaleciendo su posición en el mercado.

### *7.6.2 Innovación Continua*

La nube facilita la innovación constante, permitiendo a SLB desarrollar y probar nuevas soluciones sin las limitaciones de una infraestructura local. La flexibilidad y la capacidad de

experimentar con diferentes tecnologías y metodologías fomentan un entorno de desarrollo ágil y creativo. Esto no solo impulsa el avance tecnológico de SLB, sino que también asegura que la empresa se mantenga a la vanguardia del mercado, liderando en innovación y desarrollo de servicios.

### *7.6.3 Mejora en la Toma de Decisiones*

Contar con datos precisos y modelos predictivos avanzados mejora significativamente la toma de decisiones estratégicas en SLB. La capacidad de analizar datos en tiempo real y obtener insights precisos permite a los líderes de SLB tomar decisiones más informadas y efectivas. Por ejemplo, el uso de modelos de ML para predecir fallos en equipos críticos puede prevenir interrupciones operativas, optimizando la eficiencia y reduciendo costos.

En este capítulo se han analizado los beneficios clave que la adopción de una arquitectura en la nube puede aportar a SLB. Se destacaron aspectos como la eficiencia operativa, la escalabilidad, la seguridad mejorada y el impacto positivo en los modelos de ML. La migración a la nube no solo optimiza los recursos y procesos internos de SLB, sino que también fortalece su capacidad para enfrentar los desafíos del mercado actual.

Para la implementación efectiva de la arquitectura en la nube, se recomienda a SLB considerar varios aspectos clave. Es fundamental invertir en la, la formación y capacitación del personal en el uso de nuevas tecnologías y herramientas en la nube, son cruciales para para aprovechar al máximo las herramientas y tecnologías en la nube. Además, se sugiere establecer mantener una estrategia de seguridad robusta y actualizada, aprovechando las capacidades avanzadas de los proveedores de servicios en la nube, garantizará la protección continua de los datos sensibles y la integridad del sistema; una estrategia robusta de gestión de datos y seguridad, es esencial para proteger la información sensible, integrando prácticas de monitorización continua y optimización de costos, que es crucial

para maximizar los beneficios de la nube a largo plazo. Con estas recomendaciones, SLB podrá aprovechar al máximo los beneficios de la arquitectura en la nube y asegurar un crecimiento sostenible y competitivo en el mercado, con estas medidas, SLB puede maximizar los beneficios de la nube, asegurando una transición exitosa y una operación optimizada a largo plazo.

La adopción de una arquitectura en la nube posiciona a SLB de manera estratégica para enfrentar futuros retos y aprovechar nuevas oportunidades en la era digital. Al estar basada en una infraestructura flexible y escalable, SLB puede adaptarse rápidamente a cambios en el entorno empresarial y tecnológico, Además, la capacidad de aprovechar los últimos avances en tecnologías de ML y big data asegura que SLB continuará mejorando la precisión y efectividad de sus modelos, impulsando un ciclo continuo de mejora e innovación, que aseguran que SLB pueda mantenerse a la vanguardia y como referente dentro de la industria.

Tabla 1. Resumen de apartados capítulo 7

| Título | Subtítulo | Resumen |
|---|---|---|
| 7.1 Introducción | Contexto del Capítulo | Importancia de adoptar una arquitectura en la nube en la era digital y objetivos del capítulo: analizar beneficios específicos para SLB. |
| 7.2 Eficiencia Operativa | 7.2.1 Reducción de Costos | Comparación de costos entre infraestructura local y en la nube, con ahorros en mantenimiento y actualización de hardware. |

| | | |
|---|---|---|
| | 7.2.2 Optimización de Recursos | Uso eficiente de recursos computacionales y automatización de procesos y tareas rutinarias. |
| | 7.2.3 Flexibilidad y Adaptabilidad | Capacidad de ajustar recursos según demanda, permitiendo a SLB adaptarse rápidamente a cambios en el mercado. |
| | 7.2.4 Automatización de Procesos | Descripción de cómo los servicios en la nube permiten la automatización de tareas repetitivas, reduciendo errores humanos y aumentando la productividad. |
| 7.3 Escalabilidad | 7.3.1 Escalabilidad Vertical y Horizontal | Explicación de ambos tipos de escalabilidad y ejemplos de cómo SLB puede beneficiarse de cada tipo. |
| | 7.3.2 Gestión de Picos de Trabajo | Capacidad de manejar cargas de trabajo fluctuantes con casos de uso específicos en SLB. |
| | 7.3.3 Expansión Global | Facilidades para desplegar servicios en múltiples regiones geográficas y su impacto en la operación global de SLB. |
| 7.4 Seguridad | 7.4.1 Protocolos de Seguridad Avanzados | Descripción de medidas de seguridad ofrecidas por proveedores de nube y cómo protegen los datos de SLB. |
| | 7.4.2 Gestión de Accesos y Permisos | Implementación de controles de acceso detallados y ventajas de una gestión centralizada de accesos. |
| | 7.4.3 Resiliencia y Recuperación ante Desastres | Opciones de recuperación ante desastres proporcionadas por la nube y beneficios para la continuidad del negocio de SLB. |

| | | |
|---|---|---|
| | 7.4.4 Protección de Datos | Medidas de seguridad avanzadas ofrecidas por proveedores de servicios en la nube y ejemplos de cómo SLB puede proteger datos sensibles. |
| | 7.4.5 Cumplimiento y Normativas | Cómo los servicios en la nube ayudan a cumplir regulaciones y normativas de la industria, con casos específicos de cumplimiento en SLB. |
| 7.5 Impacto en Modelos de ML | 7.5.1 Mejora en la Precisión | Cómo la nube permite el uso de data sets más grandes y variados, mejorando la precisión de los modelos de SLB. |
| | 7.5.2 Generalización de Resultados | Capacidad de entrenar modelos más robustos y generalizables, impactando la aplicación práctica de los modelos. |
| | 7.5.3 Reducción de Tiempos de Entrenamiento | Uso de recursos computacionales avanzados para acelerar el entrenamiento, con ejemplos de reducciones de tiempo y costos en SLB. |
| | 7.5.4 Actualización y Reentrenamiento de Modelos | Facilidad de actualizar y reentrenar modelos de ML en la nube, mejorando su adaptabilidad y relevancia a lo largo del tiempo. |
| 7.6 Casos de Uso Específicos en SLB | 7.6.1 Monitoreo de Seguridad Laboral | Beneficios directos en la detección de incumplimientos de seguridad y mejoras en la respuesta a incidentes. |
| | 7.6.2 Gestión de Datos y Análisis Predictivo | Mejora en la gestión de grandes volúmenes de datos y su análisis predictivo. |
| 7.7 Impacto en SLB | 7.7.1 Aumento de la Competitividad | Análisis de cómo la adopción de una arquitectura en la nube puede dar una ventaja competitiva a SLB, sobresaliendo frente a sus competidores. |

| | | |
|---|---|---|
| | 7.7.2 Innovación Continua | Cómo la nube facilita la innovación y el desarrollo de nuevas soluciones, impactando la capacidad de SLB para liderar en el mercado. |
| | 7.7.3 Mejora en la Toma de Decisiones | Beneficios de contar con datos precisos y modelos predictivos avanzados para la toma de decisiones estratégicas en SLB. |
| Conclusión | Resumen de los Beneficios | Recapitulación de los principales beneficios analizados en el capítulo. |
| | Visión a Futuro | Reflexión sobre cómo la adopción de una arquitectura en la nube posiciona a SLB para futuros retos y oportunidades. |
| | Recomendaciones Finales | Sugerencias para la implementación efectiva de la arquitectura en la nube y maximización de sus beneficios. |

*Fuente: Autor*

# 8 RESULTADOS

En este capítulo se realiza el análisis de los resultados obtenidos durante la investigación para la propuesta de la implementación de la estrategia de despliegue del sistema de monitoreo y seguridad laboral basado en inteligencia artificial en Schlumberger (SLB). Se examinan los datos generados a partir del entrenamiento y despliegue de modelos de detección de objetos basados en IA y ML, así como la integración de estos sistemas en la infraestructura de la nube.

## 8.1 PRUEBAS REALIZADAS DURANTE EL PROCESO DE ENTRENAMIENTO DEL MODELO ML PARA SISTEMA DE MONITOREO Y SEGURIDAD LABORAL

### *8.1.1 Resultados de entrenamiento de modelo de ML usando YOLOv5 en servidor local*

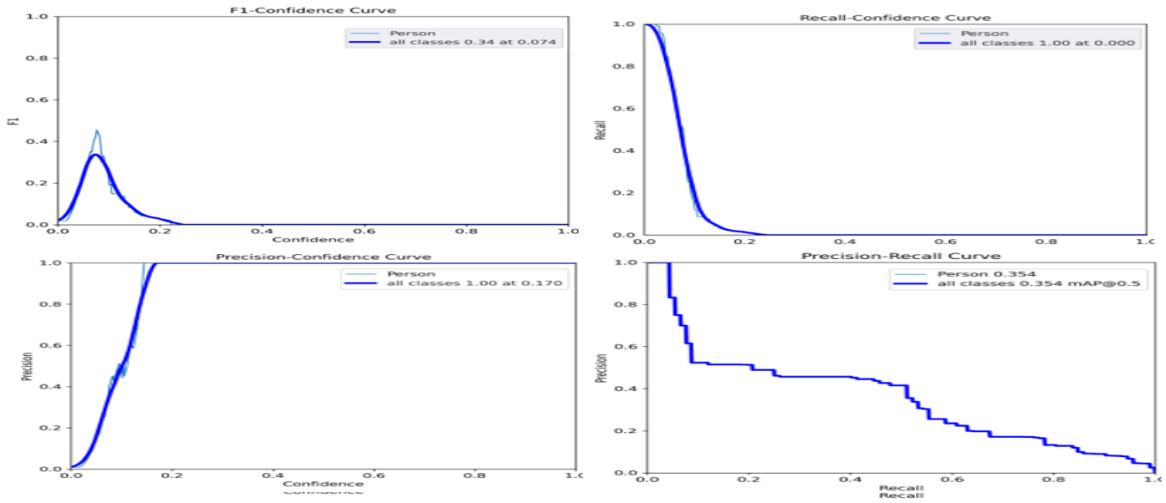


Figura 22. Graficas de confianza y precision del modelo en el reconocimiento de objetos

*Fuente: Resultados de entrenamiento modelo de ML usando YOLOv5 en servidor local*

- **Gráfica de Pérdida (Loss):** La gráfica de pérdida muestra cómo disminuye la pérdida de datos del modelo a medida que avanza el entrenamiento. En un entrenamiento ideal, se esperaría ver una disminución continua y significativa de la pérdida. Si la pérdida se estabiliza en valores altos o empieza a aumentar nuevamente, puede ser un signo de sobreajuste(overfitting), observar el patrón de la curva ayudará a determinar si el modelo necesita más entrenamiento o ajustes en los hiperparámetros. En este caso, la gráfica muestra una tendencia decreciente en la pérdida, lo que indica que el modelo está ajustándose rápidamente a los datos de entrenamiento.

- **Gráfica de Precisión (Precisión):** La precisión mide la proporción de verdaderos positivos entre los ejemplos etiquetados como positivos dentro del data set, una curva de precisión que aumenta y se estabiliza en un valor alto es deseable, pero, bajos valores de precisión pueden sugerir problemas con la calidad de la data set o la necesidad de ajustar los hiperparámetros para que aumente la precisión de reconocimiento del modelo. En este caso, la precisión aumenta y luego se estabiliza lo que significa que el data set y los datos que está manejando del modelo de ML trabajan de manera óptima dentro de su entrenamiento.

- **Gráfica de Recall:** El recall (o sensibilidad) mide la proporción de verdaderos positivos detectados entre todos los positivos reales, esta métrica va de la mano con la precisión anteriormente mencionada. Idealmente, esta métrica debe aumentar y estabilizarse cerca de 1, un recall debajo de esta métrica sugiere que el modelo no está detectando todos los objetos, si el recall es bajo, podría mejorarse con un mejor balance de clases en el data set o ajustando los umbrales de detección. En este caso, el recall aumenta de forma similar a la precisión.

- **Gráfica F1-Score:** El F1-score es la media armónica entre la precisión y el recall. Un F1-score alto y estable sugiere que el modelo tiene un buen balance entre precisión y recall. Un F1-score bajo indica problemas en el balance entre precisión y recall, lo que puede requerir ajustes en el entrenamiento para aumentar los valores tanto la precisión como el recall.

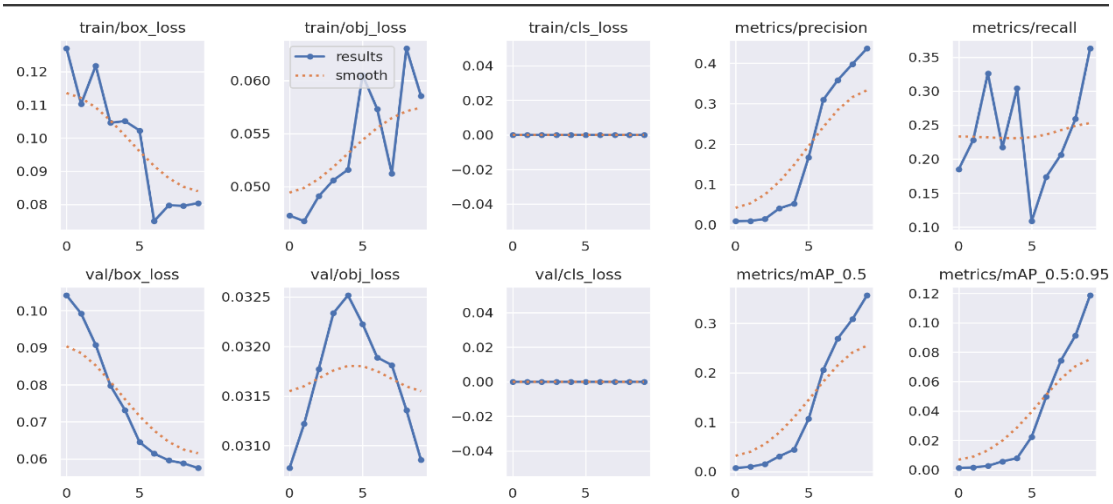


Figura 23. Gráficas estadísticas de resultados de detección de personas

*Fuente: Resultados de entrenamiento modelo de ML usando YOLOv5 en servidor local*

A continuación, se realizará el análisis de cada una de las gráficas presentadas en la figura 23, la cual se compone de 2 filas por 5 columnas. Las gráficas muestran diferentes métricas relacionadas con el proceso de entrenamiento del modelo YOLO v5.

- **Gráfica 1 (train/box_loss):** Muestra la pérdida de las cajas delimitadoras (bounding boxes) durante el entrenamiento. La pérdida de las cajas disminuye a lo largo de las épocas, lo que indica que el modelo está mejorando en la predicción de las cajas delimitadoras de los objetos, la tendencia descendente es un buen signo de que el modelo está aprendiendo a localizar los objetos correctamente. Pero no ideal como la referencia de la línea punteada.

- **Gráfica 2 (train/obj_loss):** Esta gráfica muestra una disminución en la pérdida de los objetos, aunque con cierta variabilidad, la disminución general indica que el modelo está mejorando en la detección de objetos. Sin embargo, las oscilaciones pueden sugerir inestabilidad en el proceso de entrenamiento o la necesidad de ajustar el tamaño del batch o la tasa de aprendizaje.

- **Gráfica 3 (train/cls_loss):** La pérdida de clasificación permanece constante y cercana a cero, lo cual puede indicar que el modelo no está aprendiendo adecuadamente a clasificar los objetos, esto podría ser un problema con los datos de entrenamiento o con el propio modelo.

- **Gráfica 4 (metrics/precisión):** La precisión aumenta constantemente a lo largo de las épocas, lo cual es un buen indicador de que el modelo está mejorando en la identificación correcta de objetos, que se interpreta como un buen desempeño en la predicción de verdaderos positivos y en la minimización de falsos positivos.

- **Gráfica 5 (metrics/recall):** El recall muestra un aumento significativo, pero con variabilidad. La variabilidad puede ser un signo de que el modelo aún no es completamente estable en la detección de todos los objetos. Un recall alto es deseable ya que indica que el modelo está detectando la mayoría de los objetos.

- **Gráfica 6 (val/box_loss):** La disminución de la pérdida de las cajas en el conjunto de validación es similar a la de entrenamiento, indicando que el modelo está generalizando bien y no está sobre ajustando los datos de entrenamiento.

- **Gráfica 7 (val/obj_loss):** La pérdida de objetos en validación también disminuye con menor variabilidad comparado con el entrenamiento, lo que sugiere que el modelo es más estable en datos no vistos.

- **Gráfica 8 (val/cls_loss):** Similar a la pérdida de clasificación durante el entrenamiento, esta gráfica muestra una pérdida constante cercana a cero. Esto puede confirmar problemas en la capacidad del modelo para clasificar correctamente los objetos.

- **Gráfica 9 (metrics/mAP_0.5):** El mAP a 0.5 aumenta constantemente, lo que indica que el modelo está mejorando en la precisión de las predicciones de objetos, considerando una superposición del 50% entre las predicciones y las cajas reales.

- **Gráfica 10 (metrics/mAP_0.5:0.95):** Un aumento constante en esta métrica sugiere que el modelo está mejorando su precisión en un rango más estricto de superposición, lo que resulta en un buen indicativo de que el modelo no solo predice correctamente, sino que también es preciso en la localización de los objetos para los cuales se está entrenando para identificar.

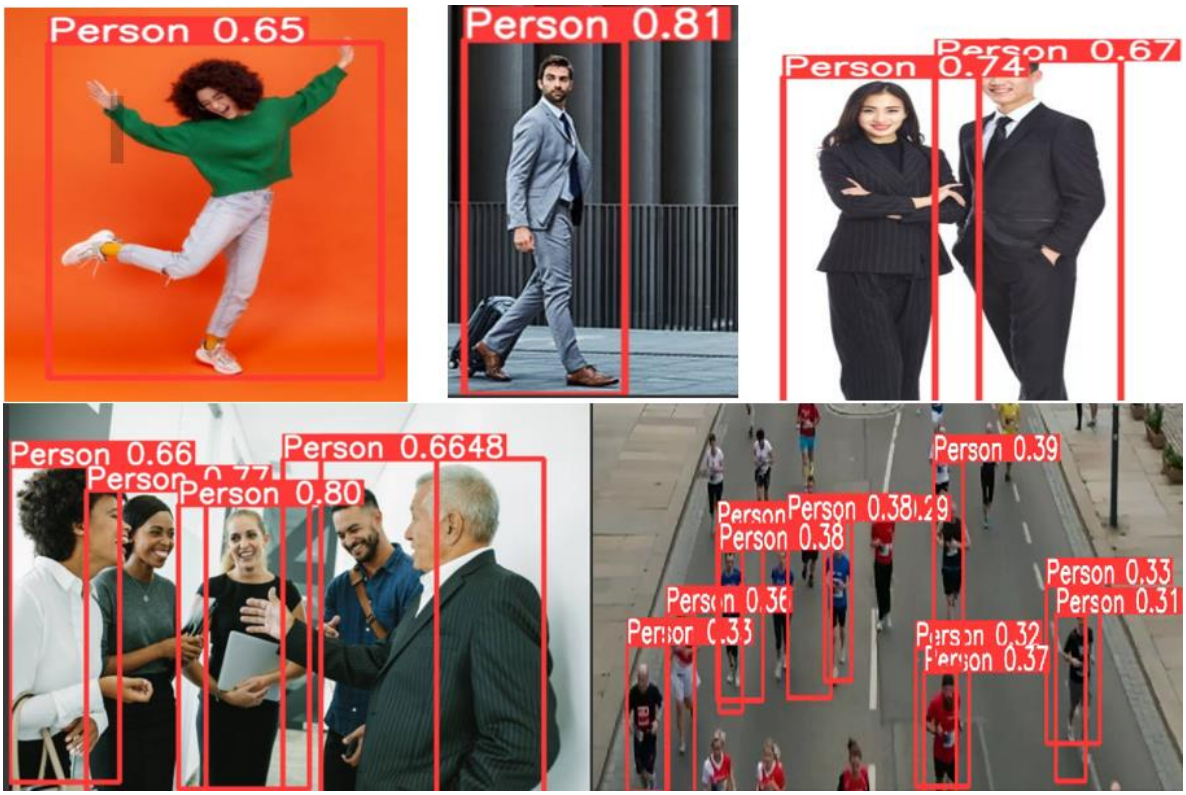


Figura 24. Resultados detección de imágenes del modelo de ML usando YOLOv5

*Fuente: Resultados de entrenamiento modelo de ML usando YOLOv5 en servidor local*

### *8.1.2 Resultados de entrenamiento de modelo de ML usando YOLOv5 en GCP*

En este apartado se presentan los resultados obtenidos al entrenar el modelo de machine learning YOLOv5 utilizando la infraestructura de GCP. Se analizan los parámetros de entrenamiento, la precisión alcanzada, y el rendimiento del modelo en términos de velocidad y eficiencia computacional. Además, se comparan estos resultados con los obtenidos en otros entornos de entrenamiento para evaluar las ventajas específicas de utilizar GCP para este propósito.

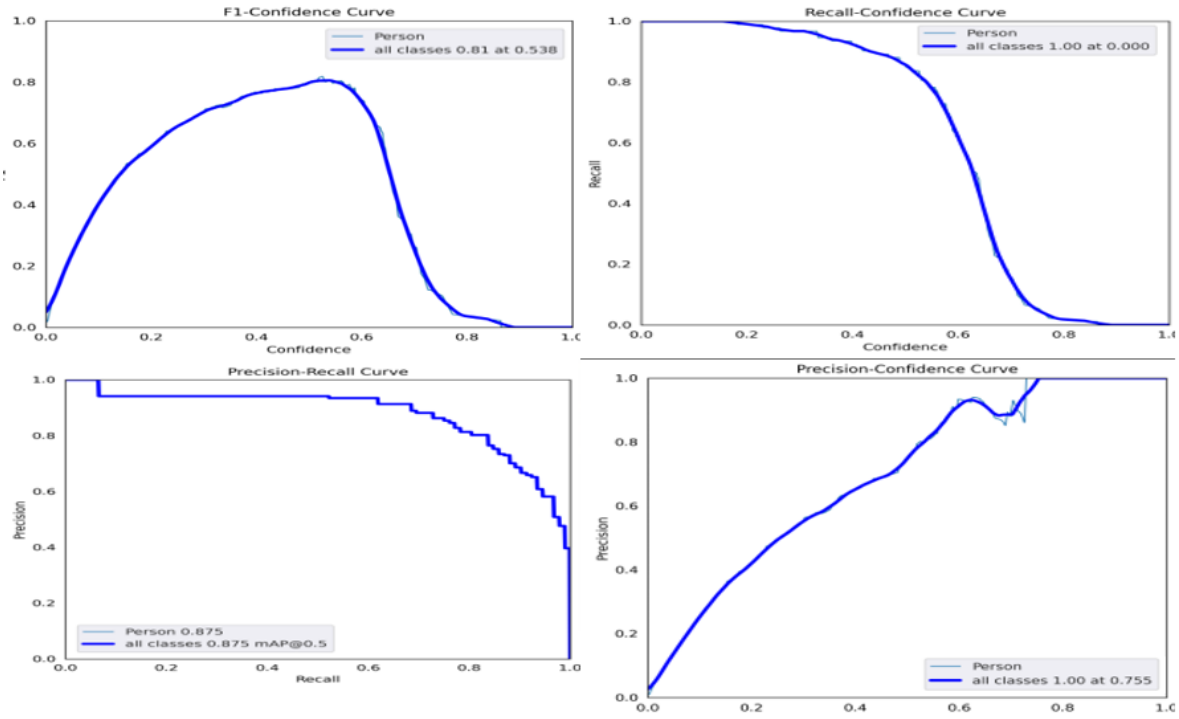


Figura 25. Graficas de confianza y precisión del modelo en el reconocimiento de objetos

*Fuente: Resultados de entrenamiento modelo de ML usando YOLOv5 GCP*

- **Gráfica de Pérdida (Loss):** La pérdida disminuye de manera más gradual, seo bserva una disminución más suave y prolongada de la pérdida en comparación con el entrenamiento anterior, gracias a un mayor tamaño de batch resulta una disminución más estable en la gráfica de perdida y la tasa de convergencia puede ser más lenta debido al aumento en el número de épocas y tamaño de batch en el entrenamiento del modelo de ML.

- **Gráfica de Precisión (Precision):** Con más épocas, la precisión debe tiene más tiempo para mejorar y estabilizarse en un valor más alto, comparando esta gráfica con la del entrenamiento anterior que fue de 50 épocas puede observarse bastantes mejoras en la estabilidad de la precisión del modelo de ML. Un mejor ajuste de

hiperparámetros puede ayudar a alcanzar valores más altos de precisión en el reconocimiento del modelo de ML.

- **Gráfica de Recall:** Similarmente, el recall evidencia mejoras al contar con más épocas de entrenamiento, permitiendo al modelo detectar más positivos verdaderos con mayor confianza. Un recall más alto y estable en comparación con el entrenamiento anterior, se muestra que el modelo está mejorando en la detección de objetos.

- **Gráfica F1-Score:** Se observa un F1-score más alto y estable comparado con el anterior indica un mejor balance entre precisión y recall, beneficiándose del mayor número de épocas y tamaño de batch indicando mejoras en el rendimiento general del modelo.


Figura 26. Gráficas estadísticas de resultados de detección de personas

*Fuente: Resultados de entrenamiento modelo de ML usando YOLOv5 GCP*

A continuación se hará el analisis de cada una de las gráficas presentadas en lafigura 26, que nuevamente se organizan en 2 filas por 5 columnas y muestran diferentes métricas relacionadas con el proceso de entrenamiento del modelo YOLO v5, pero en esta ocasión parece que con un mayor número de épocas de entrenamiento.

- **Gráfica 1 (train/box_loss):** La pérdida de las cajas disminuye notablemente a lo largo de las épocas, lo que indica una mejora continua en la capacidad del modelo para predecir las ubicaciones de las cajas delimitadoras. La reducción de la pérdida es consistente, lo que sugiere un buen ajuste del modelo.

- **Gráfica 2 (train/obj_loss):** La pérdida de los objetos también muestra una disminución con fluctuaciones menores. Esta tendencia descendente sugiere que el modelo está mejorando en la detección de objetos, aunque las oscilaciones podrían indicar la necesidad de un ajuste adicional en los hiperparámetros de entrenamiento.

- **Gráfica 3 (train/cls_loss):** La pérdida de clasificación permanece constante y cercana a cero, al igual que en la anterior serie de gráficas. Esto continúa indicando que el modelo no está mejorando en la clasificación de los objetos. Podría ser necesario revisar los datos de entrenamiento o la estructura del modelo para este aspecto en particular.

- **Gráfica 4 (metrics/precision):** La precisión aumenta de manera constante a lo largo de las épocas, lo que indica que el modelo está mejorando en la identificación correcta de objetos, minimizando los falsos positivos. Una alta precisión es un buen indicador de la capacidad del modelo para detectar correctamente los objetos.

- **Gráfica 5 (metrics/recall):** El recall muestra una tendencia ascendente con alguna variabilidad, pero en general indica una mejora en la detección de objetos, reduciendo los falsos negativos. La variabilidad podría ser debida a fluctuaciones en el entrenamiento y podría estabilizarse con un ajuste de los hiperparámetros.

- **Gráfica 6 (val/box_loss):** La disminución en la pérdida de las cajas en el conjunto de validación es similar a la de entrenamiento, lo que sugiere que el modelo está generalizando bien a datos no vistos y no está sobreajustado a los datos de entrenamiento.

- **7. Gráfica 7 (val/obj_loss):** La pérdida de objetos en validación muestra una disminución constante con menor variabilidad, lo cual es un buen signo de que el modelo es más estable y generaliza bien a nuevos datos.

- **8. Gráfica 8 (val/cls_loss):** Similar a la pérdida de clasificación durante el entrenamiento, esta gráfica muestra una pérdida constante cercana a cero. Esto refuerza la idea de posibles problemas en la capacidad del modelo para clasificar correctamente los objetos.

- **9. Gráfica 9 (metrics/mAP_0.5):** El mAP a 0.5 muestra una tendencia ascendente clara, lo que indica que el modelo está mejorando en la precisión de las predicciones de objetos. Una alta mAP a 0.5 sugiere una buena precisión con una superposición moderada entre las predicciones y las cajas reales.

- **10. Gráfica 10 (metrics/mAP_0.5:0.95):** El aumento en esta métrica sugiere una mejora continua en la precisión del modelo en un rango más estricto de

superposición. Esto es indicativo de que el modelo no solo predice correctamente, sino que también es preciso en la localización de los objetos.

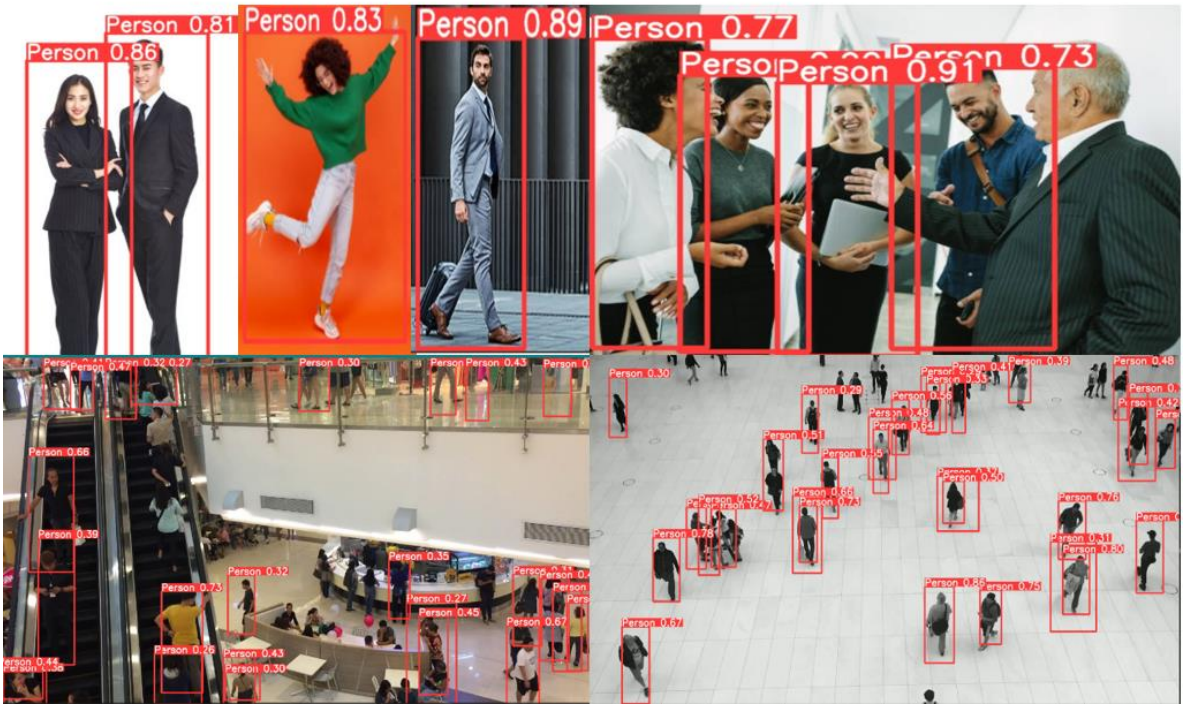


Figura 27. Gráficas de resultados de detección de personas

*Fuente: Resultados de entrenamiento modelo de ML usando YOLOv5 en GCP*

### *8.1.3 Resultados de entrenamiento de modelo de ML usando YOLOv8 en GCP*

Este apartado continúa con un análisis detallado de los resultados del entrenamiento del modelo esta vez usando la versión YOLOv8 en GCP, evidenciando métricas adicionales de rendimiento y en la evaluación de la escalabilidad del modelo en un entorno de nube. Se discuten las mejoras observadas en la precisión y eficiencia del modelo, así como los desafíos enfrentados y las soluciones implementadas durante el proceso de entrenamiento en GCP.

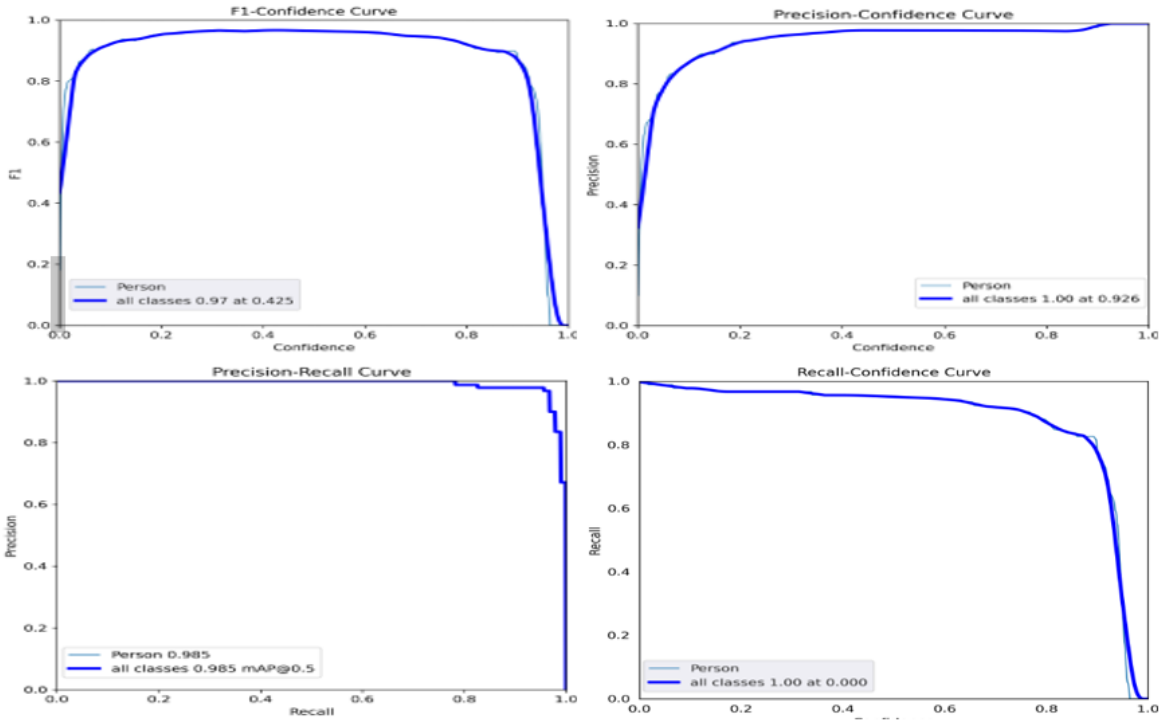


Figura 28. Graficas de confianza y precisión del modelo en el reconocimiento de objetos

*Fuente: Resultados de entrenamiento modelo de ML usando YOLOv8 en GCP*

- **Gráfica de Pérdida (Loss):** La pérdida disminuye gradualmente y se estabiliza mostrando una disminución continua a lo largo de las épocas del entrenamiento, con un posible estancamiento hacia el final si el modelo alcanza su capacidad máxima de aprendizaje, lo que significa que después de cierto número de épocas, el modelo de ML ya no necesita de más épocas para ser entrenado por que los resultados son invariantes de ahí en adelante. También se determinó que la pérdida en este entrenamiento es la más baja entre las tres configuraciones, mostrando el máximo aprovechamiento del data set.

- **Gráfica de Precisión (Precision):** La precisión empieza a estabilizarse en su valor máximo en las últimas épocas, mostrando la capacidad del modelo para generalizar correctamente tras un entrenamiento prolongado, observando una precisión mejor comparada con los de anteriores. La precisión es la más alta y estable, reflejando un entrenamiento extenso y eficiente del modelo.

- **Gráfica de Recall:** Se observa que el recall alcanza valores altos y estables, esto sugiere que el modelo está detectando la mayoría de los positivos verdaderos después de un entrenamiento exhaustivo indicando detección efectiva del modelo de ML. El valor de este recall demuestra que es el mejor entre las tres configuraciones, demostrando una excelente capacidad de detección e indicando que el modelo está detectando casi todos los objetos relevantes.

- **Gráfica F1-Score:** El F1-score es el más alto y estable en comparación con los entrenamientos anteriores, reflejando un excelente balance entre precisión y recall debido a la mayor cantidad de datos procesados y al refinamiento del modelo, así como también la efectividad de procesamiento de la nueva generación de YOLOv10 en comparación de YOLOv5.

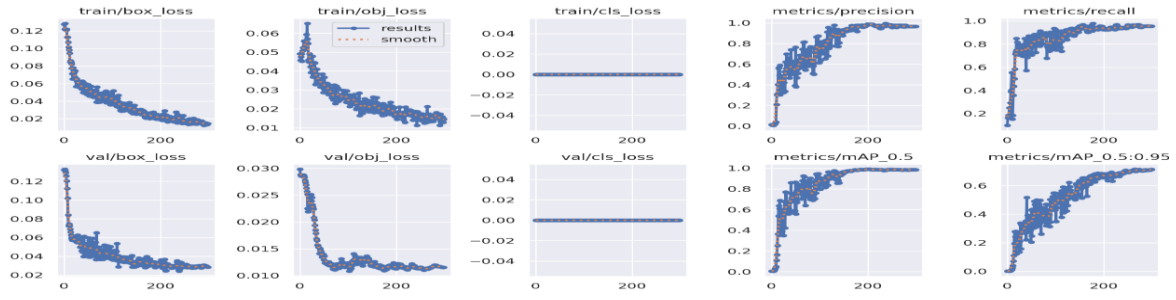


Figura 29. Gráficas estadísticas de resultados de detección de personas

*Fuente: Resultados de entrenamiento modelo de ML usando YOLOv8 en GCP*

Al a analizar cada una de las gráficas presentadas en la figura 29, se muestran los resultados de diferentes procesos de entrenamiento del modelo YOLO v5. Cada una de estas gráficas tiene una importancia particular en la evaluación del rendimiento del modelo.

- **Gráfica 1: `train/box_loss`:** La pérdida de las cajas disminuye notablemente a lo largo de las épocas, mostrando una tendencia clara hacia la baja. Esto indica que el modelo mejora continuamente su capacidad para predecir las ubicaciones de las cajas delimitadoras. La reducción consistente de la pérdida sugiere un ajuste adecuado del modelo.

- **Gráfica 2: `train/obj_loss`:** La pérdida de los objetos también muestra una disminución pronunciada con algunas fluctuaciones menores. La tendencia descendente sugiere que el modelo está mejorando en la detección de objetos a lo largo del tiempo.

- **Gráfica 3: `train/cls_loss`:** La pérdida de clasificación permanece constante y cercana a cero. Este comportamiento constante indica que el modelo no está mejorando en la clasificación de los objetos. Es posible que se necesite revisar los datos de entrenamiento o la arquitectura del modelo para este aspecto específico.

- **Gráfica 4: `metrics/precision`:** La precisión aumenta de manera constante hasta estabilizarse cerca de 1.0. Esto indica que el modelo está mejorando en la identificación correcta de objetos y minimizando los falsos positivos. Una alta precisión es un buen indicador de la capacidad del modelo para detectar correctamente los objetos.

- **5. Gráfica 5: `metrics/recall`:** El recall también muestra una tendencia ascendente, estabilizándose cerca de 1.0, lo que indica una mejora continua en la detección de objetos y una reducción de los falsos negativos.

- **6. Gráfica 6: `val/box_loss:** La disminución en la pérdida de las cajas en el conjunto de validación es similar a la de entrenamiento, sugiriendo que el modelo está generalizando bien a datos no vistos y no está sobreajustado a los datos de entrenamiento.

- **7. Gráfica 7: `val/obj_loss`:** La pérdida de objetos en validación también muestra una disminución constante con menor variabilidad, lo cual es un buen signo de que el modelo es estable y generaliza bien a nuevos datos.

- **8. Gráfica 8: `val/cls_loss`:** Similar a la pérdida de clasificación durante el entrenamiento, esta gráfica muestra una pérdida constante cercana a cero. Esto refuerza la idea de que hay un problema en la capacidad del modelo para clasificar correctamente los objetos.

- **9. Gráfica 9: `metrics/mAP_0.5`:** El mAP a 0.5 muestra una tendencia ascendente clara, estabilizándose cerca de 1.0, lo que indica que el modelo está mejorando en la precisión de las predicciones de objetos.

- **10. Gráfica 10: `metrics/mAP_0.5:0.95`:** El aumento en esta métrica y su estabilización cerca de 0.8 sugiere una mejora continua en la precisión del modelo en un rango más estricto de superposición. Esto indica que el modelo no solo predice correctamente, sino que también es preciso en la localización de los objetos.


Figura 30. Resultado de deteccion de personas en imagenes

*Fuente: Resultados de entrenamiento modelo de ML usando YOLOv8*

Tras evaluar los resultados de entrenamiento del modelo de ML YOLOv8 en GCP, presentados en la sección 8.1.1, se verificó que el modelo alcanzó los niveles de precisión y confianza requeridos para la detección de personas. Posteriormente, se aplicó el modelo a análisis de video, obteniendo resultados satisfactorios. Con base en estos hallazgos, se concluye que el modelo está listo para ser implementado en sistemas de seguridad, permitiendo la detección de personas en áreas delimitadas y la generación de alertas en caso de intrusiones.

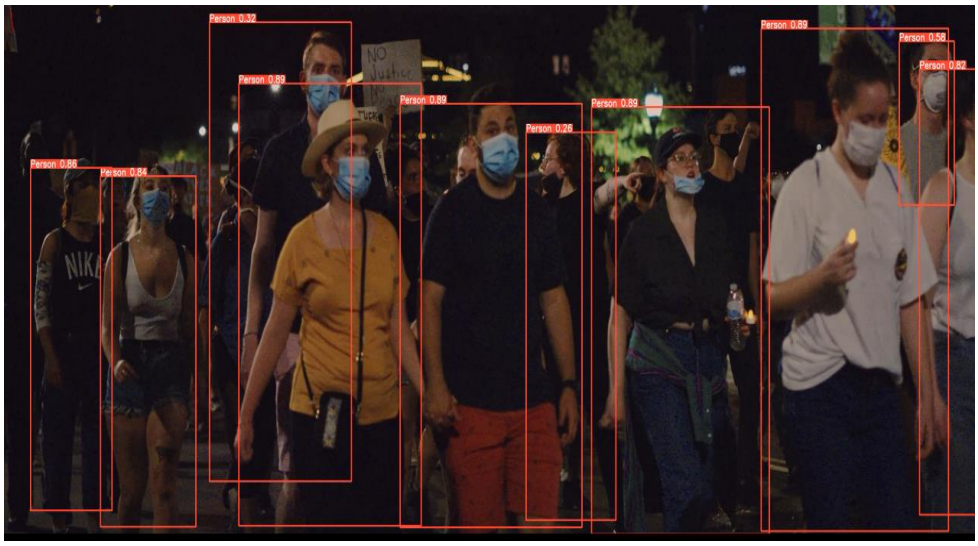

Figura 31. Resultados de detección de personas cambiando luminosidad y angulo de la camara

*Fuente: resultados de entrenamiento modelo de ML usando YOLOv8*

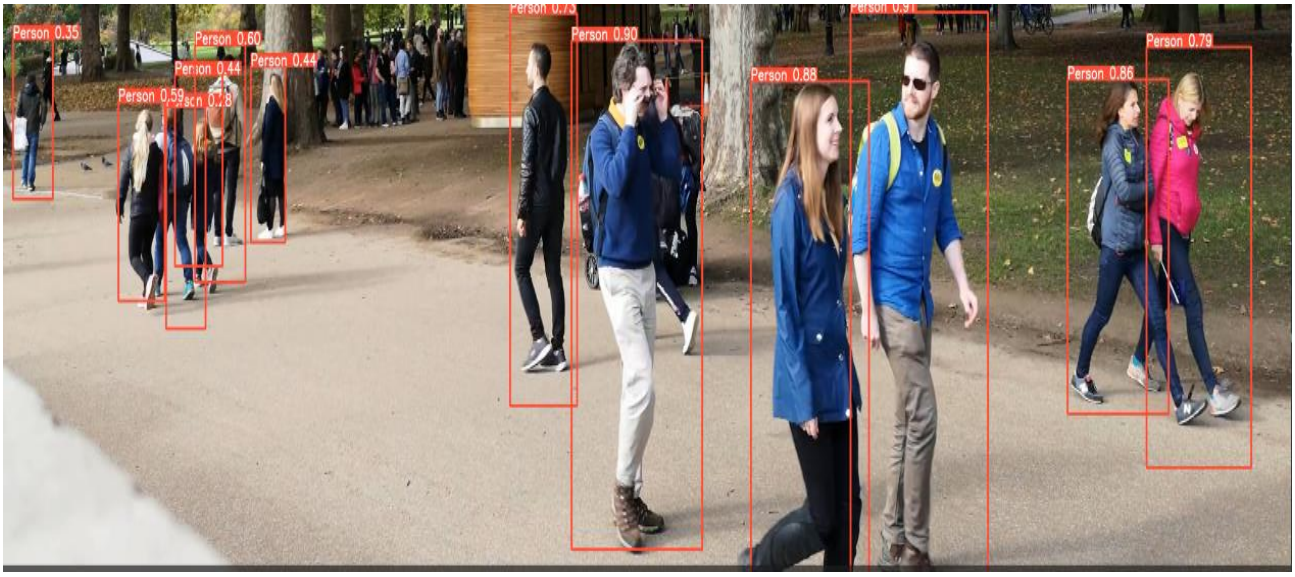

Figura 32. Resultados de detección de personas cambiando luminosidad de la camara

*Fuente: resultados de entrenamiento modelo de ML usando YOLOv8*

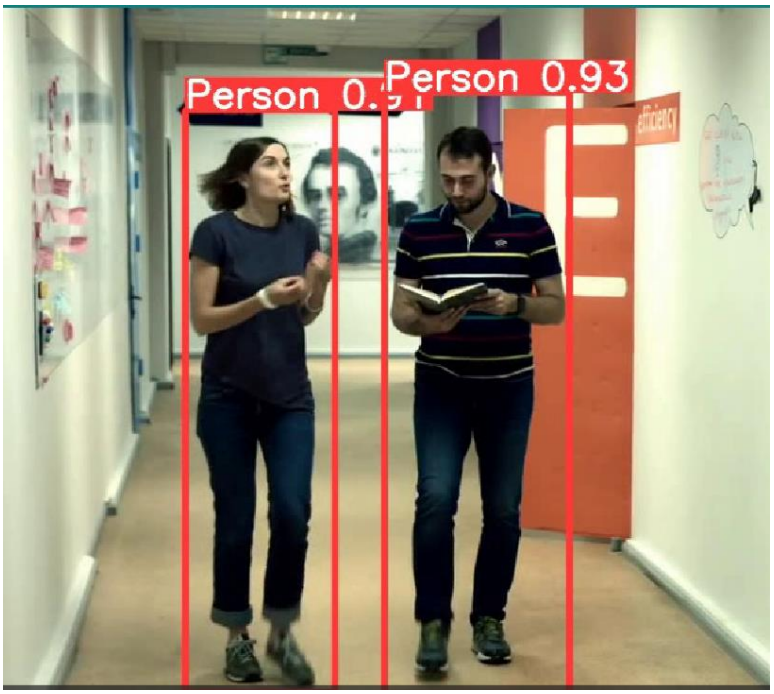


Figura 33. Resultados de detección de personas cambiando angulo de la camara

*Fuente: resultados de entrenamiento modelo de ML usando YOLOv8*

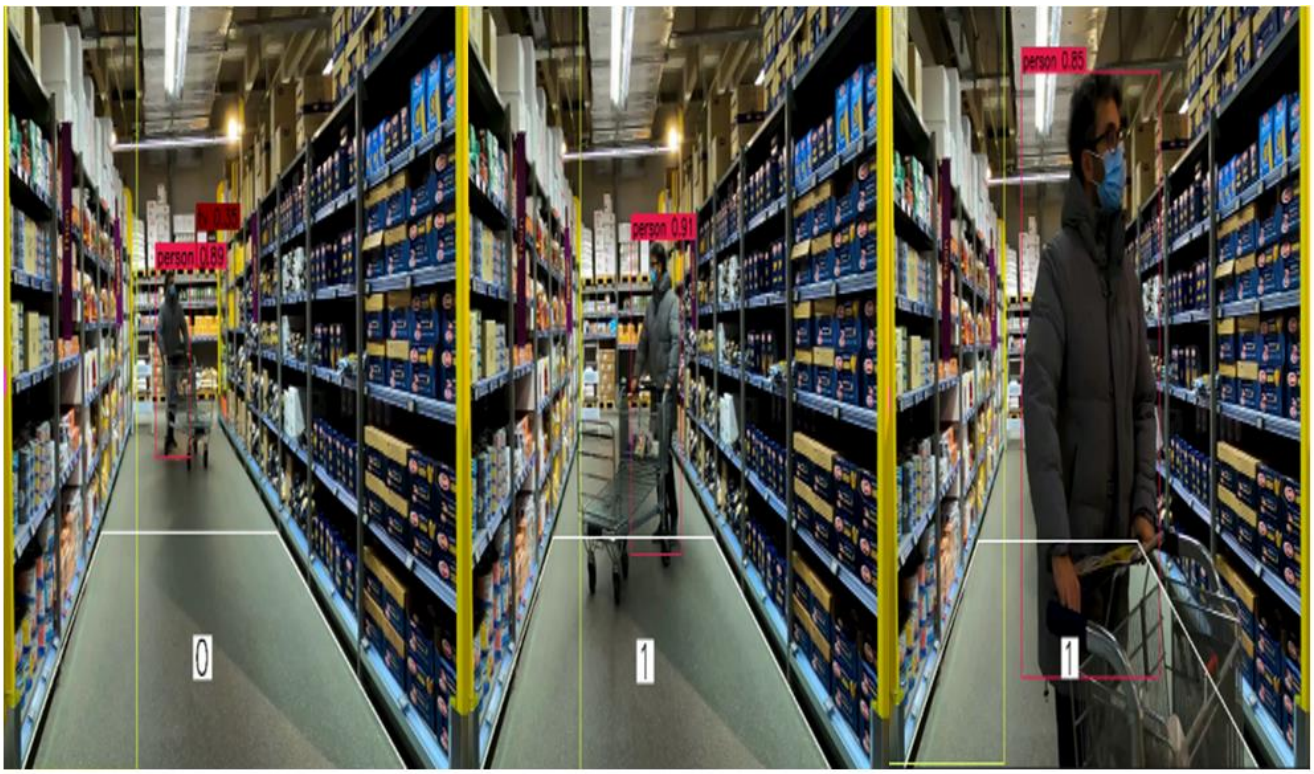


Figura 34. Resultados de detección de personas cambiando angulo de la camara y delimitando zona de reconocimiento para enviar alertas

*Fuente: resultados de entrenamiento modelo de ML usando YOLOv8*

## 8.2 ANÁLISIS DE RESULTADOS OBTENIDOS ENTRE LOS ENTRENAMIENTOS

En el apartado anterior, se analizó la eficiencia, efectividad y características de tres entornos distintos para entrenar modelos de detección de objetos utilizando YOLO en su versión small, elegido especialmente por la cantidad de datos recolectados para realizar las pruebas en cada entorno.

### *8.2.1 Eficiencia y Efectividad en el Entrenamiento de modelos de ML*

La eficiencia se refiere a la velocidad de entrenamiento y la utilización de recursos (CPU, GPU, memoria), mientras que la efectividad se relaciona con la calidad del modelo resultante, medida por métricas como mAP (mean Average Precision).

Tabla 2. Comparación de los 3 entornos en los que se entrenó el modelo de ML

| **Característica** | **Entrenamiento Local** | **GCP (YOLOv5)** | **GCP (YOLOv8)** |
|---|---|---|---|
| **Eficiencia** | | | |
| **Velocidad** | Más lenta | Más rápida | Mucho más rápida |
| **Utilización de recursos** | Moderada | Alta (compartida) | Muy alta (personalizable) |
| **Costo por hora** | Bajo | Gratuito (limitado) | Alto |
| **Efectividad** | | | |
| **mAP** | Puede variar | Generalmente bueno | Potencialmente superior (YOLOv8) |
| **Generalización** | Puede ser limitada | Buena | Excelente |
| **Otras Consideraciones** | | | |
| **Flexibilidad** | Alta | Media | Alta |
| **Escalabilidad** | Baja | Media | Alta |
| **Facilidad de uso** | Alta | Muy alta | Media |

*Fuente: Autor*

- **Entrenamiento Local:** Ofrece un alto grado de control y flexibilidad, pero su rendimiento está limitado por el hardware disponible. Es una buena opción para proyectos pequeños o para aquellos que prefieren tener un control total sobre el entorno de desarrollo. Sin embargo, puede resultar en tiempos de entrenamiento prolongados, especialmente para datasets grandes.

- **GCP con YOLOv5:** Brinda acceso gratuito a GPU´s aceleradoras, lo que lo convierte en una excelente opción para prototipar rápidamente y experimentar con diferentes configuraciones. Su interfaz basada en Jupyter Notebook facilita su uso. No obstante, tiene limitaciones en cuanto a tiempo de ejecución y recursos disponibles, lo que puede ser un obstáculo para proyectos más grandes.

- **GCP con YOLOv8:** Ofrece la máxima escalabilidad y flexibilidad, permitiendo ajustar la configuración de las instancias para adaptarse a las necesidades específicas de cada proyecto. Esto lo convierte en la opción ideal para proyectos a gran escala o aquellos que requieren un alto rendimiento. Sin embargo, su costo puede ser elevado, especialmente para proyectos a largo plazo.

Los resultados mejoran significativamente debido a la flexibilidad y las capacidades avanzadas de la nube, la integración con servicios de análisis y machine learning facilita la implementación de modelos más precisos y adaptativos para la detección de riesgos y la prevención de incidentes. Las actualizaciones y mejoras continuas en los servicios en la nube garantizan que el sistema de monitoreo y seguridad laboral se mantenga a la vanguardia de la tecnología, permitiendo una evolución constante y mejorada en la identificación de amenazas y la implementación de medidas preventivas, concluyendo que la nube no solo optimiza el rendimiento y reduce los costos, sino que también potencia la capacidad del sistema para proporcionar resultados más precisos y fiables.

# 9 CONCLUSIONES

- La propuesta de una arquitectura en la nube para el sistema de monitoreo y seguridad laboral en SLB se ha demostrado como una solución efectiva para optimizar los procesos internos de la empresa, al implementar esta estrategia, se logrará una mayor automatización y escalabilidad, permitiendo el despliegue eficiente de sistemas de monitoreo basados en IA. Esto no solo mejorará la capacidad de la empresa para reconocer personas y garantizar el uso adecuado de elementos de protección, sino que también facilitará el monitoreo de múltiples ubicaciones a través de una infraestructura unificada y flexible.

- El análisis del proceso actual de implementación del sistema de monitoreo en SLB reveló varias áreas de mejora, especialmente en términos de automatización y eficiencia. El sistema existente, aunque funcional, depende en gran medida de procesos manuales y una infraestructura física que limita su escalabilidad y capacidad de respuesta, este diagnóstico permitió identificar las debilidades y establecer un punto de partida claro para la transición hacia una solución basada en la nube.

- Las principales ventajas del sistema de monitoreo y seguridad, que se tiene actualmente, incluyen su capacidad para operar en tiempo real y su integración con las operaciones diarias de SLB. Sin embargo, se identificaron desventajas significativas como la falta de escalabilidad, la dependencia de equipos físicos costosos y la limitada capacidad de análisis de datos, estas desventajas subrayan la necesidad de adoptar una solución en la nube que pueda ofrecer mayor flexibilidad y eficiencia operativa.

- Se identificaron varias tecnologías y servicios clave que facilitan el despliegue de una arquitectura en la nube, incluyendo plataformas como GCP. Esta plataforma ofrece servicios específicos para almacenamiento, procesamiento y despliegue de modelos de ML, proporcionando una infraestructura robusta y escalable para el sistema de monitoreo y seguridad laboral de SLB.

- El despliegue de una arquitectura en la nube ofrece múltiples beneficios, incluyendo una mayor eficiencia operativa, escalabilidad y mejoras en la seguridad de los datos. Además, la utilización de modelos de ML más avanzados ha mejorado significativamente la precisión y generalización de los resultados, permitiendo un monitoreo más efectivo y una respuesta rápida ante incidentes de seguridad laboral.

- La estrategia propuesta para el despliegue de una arquitectura en la nube en SLB incluye la adopción de servicios avanzados de IA y ML, la integración de tecnologías de orquestación y automatización, y la implementación de medidas de seguridad robustas. Esta estrategia no solo optimiza el proceso de implementación, sino que también asegura la escalabilidad y sostenibilidad del sistema de monitoreo y seguridad laboral a largo plazo para SLB.

- La elección del entorno de entrenamiento dependerá de varios factores, como el presupuesto, la complejidad del modelo, el tamaño del conjunto de datos y los requisitos de tiempo. Por lo que hay que tener las siguientes consideraciones:

    - Si el presupuesto es limitado y se busca una solución rápida pero con resultados menos óptimos, el entrenamiento local puede ser una opción.

- Para obtener resultados de alta calidad a un costo razonable, Google Colab es una excelente opción para empezar.

- Si se requiere la máxima precisión y flexibilidad, Google GCP es la mejor opción, aunque puede ser más costosa.

# 10 REFERENCIAS


[1] “Compute Engine | Google Cloud.” Accessed: Jun. 08, 2024. [Online]. Available: https://cloud.google.com/products/compute?hl=es_419

[2] “Cloud Storage | Google Cloud.” Accessed: Jun. 08, 2024. [Online]. Available: https://cloud.google.com/storage

[3] “Google Kubernetes Engine (GKE) | Google Cloud.” Accessed: Jun. 08, 2024. [Online]. Available: https://cloud.google.com/kubernetes-engine?hl=es_419

[4] “¿Qué es Pub/Sub? | Pub/Sub Documentation | Google Cloud.” Accessed: Jun. 08, 2024. [Online]. Available: https://cloud.google.com/pubsub/docs/overview?hl=es-419

[5] “Servicios de informática en la nube | Microsoft Azure.” Accessed: Jul. 22, 2024. [Online]. Available: https://azure.microsoft.com/es-mx

[6] “¿Qué son los servidores en la nube?: Explicación sobre los servidores en la nube: AWS.” Accessed: May 21, 2024. [Online]. Available: https://aws.amazon.com/es/what-is/cloud-server/

[7] “Modelos de servicio en la nube | Tipos de cloud computing | AWS.” Accessed: May 21, 2024. [Online]. Available: https://aws.amazon.com/es/types-of-cloud-computing/

[8] “Documentation | Apache Airflow.” Accessed: Jun. 08, 2024. [Online]. Available: https://airflow.apache.org/docs/

[9] “Jenkins.” Accessed: Jun. 19, 2024. [Online]. Available: https://www.jenkins.io/

[10] “Terraform by HashiCorp.” Accessed: Jun. 08, 2024. [Online]. Available: https://www.terraform.io/

[11] “Identity and Access Management | IAM | Google Cloud.” Accessed: Jul. 22, 2024. [Online]. Available: https://cloud.google.com/security/products/iam?hl=es-419

[12] “Identity-Aware Proxy (IAP) | Google Cloud.” Accessed: Jul. 22, 2024. [Online]. Available: https://cloud.google.com/security/products/iap?hl=es_419

[13] “6 Regularización | Machine Learning: Teoría y Práctica.” Accessed: May 21, 2024. [Online]. Available: https://bookdown.org/victor_morales/TecnicasML/regularizaci%C3%B3n.html

[14] “¿Qué es el aprendizaje supervisado? | IBM.” Accessed: May 21, 2024. [Online]. Available: https://www.ibm.com/mx-es/topics/supervised-learning

[15] “¿Qué es el aprendizaje no supervisado? | IBM.” Accessed: May 21, 2024. [Online]. Available: https://www.ibm.com/mx-es/topics/unsupervised-learning

[16] “Aprendizaje por refuerzo y optimización - Expertos en IIC.” Accessed: May 21, 2024. [Online]. Available: https://www.iic.uam.es/inteligencia-artificial/aprendizaje-por-refuerzo/

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 779–788, Jun. 2015, doi: 10.1109/CVPR.2016.91.

[18] J. S. D. R. G. A. F. Redmon, “(YOLO) You Only Look Once,” *Cvpr*, vol. 2016-December, pp. 779–788, Dec. 2016, doi: 10.1109/CVPR.2016.91.

[19] C. Ning, H. Zhou, Y. Song, and J. Tang, “Inception Single Shot MultiBox Detector for object detection,” *2017 IEEE International Conference on Multimedia and Expo Workshops, ICMEW 2017*, pp. 549–554, Sep. 2017, doi: 10.1109/ICMEW.2017.8026312.

[20] R. I. H. Abushahma, M. A. M. Ali, O. I. Al-Sanjary, and N. M. Tahir, “Region-based convolutional neural network as object detection in images,” *Proceeding - 2019 IEEE*

*7th Conference on Systems, Process and Control, ICSPC 2019*, pp. 264–268, Dec. 2019, doi: 10.1109/ICSPC47137.2019.9068011.

[21] B. S. Reddy, A. Mallikarjuna Reddy, M. H. D. S. Sradda, T. Mounika, S. Mounika, and K. Meghana, “A Comparative Study on Object Detection Using Retinanet,” *MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference*, 2022, doi: 10.1109/MYSURUCON55714.2022.9972742.

[22] N. G. Camacho, “Unlocking the Potential of AI/ML in DevSecOps: Effective Strategies and Optimal Practices,” *Journal of Artificial Intelligence General science (JAIGS) ISSN:3006-4023*, vol. 2, no. 1, pp. 79–89, Mar. 2024, doi: 10.60087/JAIGS.V2I1.P89.

[23] “The Role Of Machine Learning In Autonomous Vehicles | AVs-report – Weights & Biases.” Accessed: Jul. 23, 2024. [Online]. Available: https://wandb.ai/ivangoncharov/AVs-report/reports/The-Role-Of-Machine-Learning-In-Autonomous-Vehicles--VmlldzoyNTExMDE3

[24] “The Role of Machine Learning in Health Care.” Accessed: Jul. 23, 2024. [Online]. Available: https://corporatelearning.hms.harvard.edu/blog/machine-learning-medicine-101-health-care-executives

[25] “AI in Retail: How to Use Machine Learning in eCommerce | by Sandra Parker | Medium.” Accessed: Jul. 23, 2024. [Online]. Available: https://sandra-parker.medium.com/ai-in-retail-how-to-use-machine-learning-in-ecommerce-643ef10e24fd

[26] P. Tripathi, N. Kumar, M. Rai, P. K. Shukla, and K. N. Verma, “Applications of Machine Learning in Agriculture,” *Smart Village Infrastructure and Sustainable Rural Communities*, pp. 99–118, Jan. 2023, doi: 10.4018/978-1-6684-6418-2.CH006.

[27] “How Machine Learning Will Transform Supply Chain Management.” Accessed: Jul. 23, 2024. [Online]. Available: https://hbr.org/2024/03/how-machine-learning-will-transform-supply-chain-management

[28] S. Angra, B. Sharma, and K. D. Sharma, “Amalgamation of Virtual Reality, Augmented Reality and Machine Learning: A Review,” *2022 2nd International*

*Conference on Advance Computing and Innovative Technologies in Engineering, ICACITE 2022*, pp. 2601–2604, 2022, doi: 10.1109/ICACITE53722.2022.9823716.

[29] “Batch vs. Real-Time Machine Learning: Understanding the Processing Powerhouse and the Speedy Predictor | by Sanjay Kumar | Medium.” Accessed: Jul. 23, 2024. [Online]. Available: https://medium.com/@Sanjaynk7907/batch-vs-real-time-machine-learning-understanding-the-processing-powerhouse-and-the-speedy-bb7f962b9be6

[30] “Scalability in Machine Learning Systems: Challenges, Strategies, and Best Practices | by amirsina torfi | Medium.” Accessed: Jul. 23, 2024. [Online]. Available: https://medium.com/@amirsina.torfi/scalability-in-machine-learning-systems-challenges-strategies-and-best-practices-231cc2fb2889

[31] “Understanding the Scope and Limitations of Machine Learning | by Pulkit | Medium.” Accessed: Jul. 23, 2024. [Online]. Available: https://medium.com/@itspulkit37/understanding-the-scope-and-limitations-of-machine-learning-d5e2c734b68e

[32] “División de datos - Amazon Machine Learning.” Accessed: Jul. 23, 2024. [Online]. Available: https://docs.aws.amazon.com/es_es/machine-learning/latest/dg/splitting-types.html

[33] “Data Augmentation – Towards Data Science.” Accessed: May 21, 2024. [Online]. Available: https://towardsdatascience.com/tagged/data-augmentation

[34] “Data augmentation | TensorFlow Core.” Accessed: May 21, 2024. [Online]. Available: https://www.tensorflow.org/tutorials/images/data_augmentation

[35] “What is transfer learning? | IBM.” Accessed: May 21, 2024. [Online]. Available: https://www.ibm.com/topics/transfer-learning?mhsrc=ibmsearch_a&mhq=transfer%20learning

[36] “What is Fine-Tuning? | IBM.” Accessed: May 21, 2024. [Online]. Available: https://www.ibm.com/topics/fine-tuning

[37] “¿Cuál es la diferencia entre PaaS, IaaS y SaaS? | Google Cloud | Google Cloud.” Accessed: May 27, 2024. [Online]. Available: https://cloud.google.com/learn/paas-vs-iaas-vs-saas?hl=es-419

[38] “Making AI helpful for everyone - Google AI – Google AI.” Accessed: Jun. 08, 2024. [Online]. Available: https://ai.google/

[39] “Cloud Functions | Google Cloud.” Accessed: Jun. 08, 2024. [Online]. Available: https://cloud.google.com/functions?hl=es_419

[40] “Cloud SQL para MySQL, PostgreSQL y SQL Server | Google Cloud.” Accessed: Jun. 08, 2024. [Online]. Available: https://cloud.google.com/sql?hl=es_419

[41] “BigQuery: Almacén de datos empresarial | Google Cloud.” Accessed: Jun. 08, 2024. [Online]. Available: https://cloud.google.com/bigquery?hl=es_419

[42] “API Documentation | TensorFlow v2.16.1.” Accessed: Jun. 08, 2024. [Online]. Available: https://www.tensorflow.org/api_docs

[43] “FAQ’S — PyTorch/Serve master documentation.” Accessed: Jun. 08, 2024. [Online]. Available: https://pytorch.org/serve/FAQs.html

[44] “COCO - Common Objects in Context.” Accessed: Jun. 08, 2024. [Online]. Available: https://cocodataset.org/#home

[45] “Ultralytics · GitHub.” Accessed: Jun. 08, 2024. [Online]. Available: https://github.com/ultralytics

[46] “CVAT.” Accessed: Jun. 08, 2024. [Online]. Available: https://www.cvat.ai/

[47] “Roboflow: Computer vision tools for developers and enterprises.” Accessed: Jun. 08, 2024. [Online]. Available: https://roboflow.com/

[48] “Cloud AutoML Documentation | Google Cloud.” Accessed: Jun. 08, 2024. [Online]. Available: https://cloud.google.com/automl/docs?hl=es-419

[49] “Hyperparameter optimization - Wikipedia.” Accessed: Jun. 08, 2024. [Online]. Available: https://en.wikipedia.org/wiki/Hyperparameter_optimization

[50] “Albumentations: fast and flexible image augmentations.” Accessed: Jun. 19, 2024. [Online]. Available: https://albumentations.ai/

[51] “Apache Kafka.” Accessed: Jun. 19, 2024. [Online]. Available: https://kafka.apache.org/

[52] “Apache Spark™ - Unified Engine for large-scale data analytics.” Accessed: Jun. 19, 2024. [Online]. Available: https://spark.apache.org/

[53] “Apache Hadoop.” Accessed: Jun. 08, 2024. [Online]. Available: https://hadoop.apache.org/

[54] “Vault by HashiCorp.” Accessed: Jun. 19, 2024. [Online]. Available: https://www.vaultproject.io/

[55] “¿Cómo funciona SSL? | Certificados SSL y TLS | Cloudflare.” Accessed: Jun. 08, 2024. [Online]. Available: https://www.cloudflare.com/es-es/learning/ssl/how-does-ssl-work/

[56] “Prometheus - Monitoring system & time series database.” Accessed: Jun. 08, 2024. [Online]. Available: https://prometheus.io/

[57] “ELK Stack: Elasticsearch, Kibana, Beats y Logstash | Elastic.” Accessed: Jun. 19, 2024. [Online]. Available: https://www.elastic.co/es/elastic-stack

[58] “Docker: Accelerated Container Application Development.” Accessed: Jun. 19, 2024. [Online]. Available: https://www.docker.com/

[59] “Docker: Accelerated Container Application Development.” Accessed: Jun. 08, 2024. [Online]. Available: https://www.docker.com/

[60] “What is Jenkins? Why Use Continuous Integration (CI) Tool?” Accessed: Jun. 08, 2024. [Online]. Available: https://www.guru99.com/jenkin-continuous-integration.html

[61] “Overview - Spark 3.5.1 Documentation.” Accessed: Jun. 08, 2024. [Online]. Available: https://spark.apache.org/docs/3.5.1/

## 11 LISTA DE FIGURAS

# 12 LISTA DE TABLAS