UNIVERSIDAD NACIONAL DE MOQUEGUA

FACULTAD DE INGENIERIA Y ARQUITECTURA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

INFORME DE EXAMEN PROFESIONAL

IMPLEMENTACIÓN DE LECTURA DE DOCUMENTOS DIGITALES CON MACHINE LEARNING EN LA EMPRESA CLI GESTIONES ADUANERAS S.A.

PRESENTADO POR:

BACH. DARWIN DANIEL ENRIQUE APAZA FLORES

PARA OPTAR EL TITULO PROFESIONAL DE:

INGENIERO DE SISTEMAS E INFORMATICA

MOQUEGUA – PERÚ

2022

UNIVERSIDAD NACIONAL DE MOQUEGUA

FACULTAD DE INGENIERIA Y ARQUITECTURA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

INFORME DE EXAMEN PROFESIONAL

IMPLEMENTACIÓN DE LECTURA DE DOCUMENTOS DIGITALES CON MACHINE LEARNING EN LA EMPRESA CLI GESTIONES ADUANERAS S.A.

PRESENTADO POR:

BACH. DARWIN DANIEL ENRIQUE APAZA FLORES

PARA OPTAR EL TITULO PROFESIONAL DE:

INGENIERO DE SISTEMAS E INFORMATICA

_____

BACH. DARWIN DANIEL ENRIQUE APAZA FLORES

_____

MSC. VICTOR YANA MAMANI
ASESOR

MOQUEGUA - PERÚ

2022

UNAM

ESCUELA PROFESIONAL DE
INGENIERÍA DE SISTEMAS E INFORMÁTICA

Folio Nº 017

ACTA DE TITULACIÓN

I. Modalidad:
Examen Profesional

II. Documento de aprobación:
Resolución de Comisión Organizadora Nº 636-2022-UNAM
Fecha: 19 agosto de 2022

III. Título de Informe de Examen Profesional:
Implementación de lectura de documentos digitales con machine Learning en la Empresa CUI Gestiones Aduaneras S.A.

Bachiller: Darwin Daniel Enrique Apaza Flores
Asesor Principal: M.Sc. Victor Yana Mamani
Co-Asesor:

IV. Se dictamina con las siguientes notas:

Examen Escrito: 14.40 Entrevista: 2.80 Promedio: 17

Promedio en letras: Diecisiete por lo que: Sí X No

Opta el Título Profesional de INGENIERO DE SISTEMAS E INFORMÁTICA.

Siendo las 11:00 horas, del 18 de octubre del año 2022 se procede con la firma en conformidad de lo dictaminado.

M.Sc. Hugo Euler Tito Chura
Presidente del Comité de Grados y Títulos
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

M.Sc. Victor Yana Mamani
Director
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

M.Sc. Victor Yana Mamani
Asesor
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

Bach. Darwin Apaza Flores

DEDICATORIA

A mis padres Nicolasa y Enrique, por los valores y principios que inculcaron en mí, por sus esfuerzos por lograr verme hecho un profesional y por apoyarme en todo momento.

AGRADECIMIENTO

A Dios, por escuchar mis oraciones en los momentos cuando creía desistir e iluminar mi camino cada día.

A la Universidad Nacional de Moquegua por compartir conmigo sus instalaciones durante mis años de estudio.

A mi Escuela Profesional de Ingeniería de Sistemas E Informática y docentes, por compartirme sus conocimientos, experiencias y buenos consejos.

A mis padres por el apoyo brindado cuando decidí estudiar esta carrera y siempre confiar en mí.

A mi novia por su apoyo moral incondicional durante todos mis años de estudio de pregrado, compartiendo sus opiniones en el desarrollo de muchos de mis trabajos académicos.

A mis compañeros y amigos, por compartirme su amistad, conocimiento y experiencias a lo largo de todos los años de estudio.

RESUMEN

La presente implementación tuvo como objetivo automatizar procesos de digitalización de datos apoyándose en técnicas de Inteligencia Artificial, como Machine Learning, sobre documentos aduaneros como Facturas Comerciales y Documentos de Embarque fundamentales para la emisión de una DAM a SUNAT.

Para el cumplimiento del objetivo de la implementación, se hicieron uso de servicios en la nube como Computer Vision y Form Recognizer de la plataforma de Microsoft Azure, las cuales funcionan sobre Inteligencia Artificial y Machine Learning para crear modelos personalizados de documentos aduaneros, permitiendo ahorros en tiempos de desarrollo e infraestructura tecnológica. Los tiempos que toma crear y entrenar modelos sobre su plataforma van desde segundos a pocos minutos y las muestras mínimas requeridas son de 5, haciendo que sea tomada como mejor opción ante otros servicios con funcionalidades similares.

Como objetivo alcanzado con esta implementación, los tiempos de atención por parte de área de Liquidación se redujeron hasta en un 75% haciendo esto que puedan atender más despachos en el mismo tiempo que tomaba atender solo uno. Así también, se lograron reducir los errores de digitalización que conllevaba a un proceso de rectificación y en algunos casos pagos de multas.

Palabras claves: Inteligencia Artificial, Machine Learning, Computer Vision, Form Recognizer, DAM.

ABSTRACT

The objective of this implementation was to automate data digitalization processes using Artificial Intelligence techniques, such as Machine Learning, on customs documents such as Commercial Invoices and Shipping Documents, which are essential for the issuance of a DAM to SUNAT.

To meet the implementation objective, cloud services such as Computer Vision and Form Recognizer from the Microsoft Azure platform were used, which work on Artificial Intelligence and Machine Learning to create customized models of customs documents, allowing savings in development times and technological infrastructure. The time it takes to create and train models on its platform ranges from seconds to a few minutes and the minimum required samples are 5, making it a better option than other services with similar functionalities.

As an objective achieved with this implementation, the service times by the Settlement area were reduced by up to 75%, allowing them to attend more dispatches in the same time it took to attend only one. Also, it was possible to reduce digitalization errors that led to a process of rectification and in some cases payment of fines.

Keywords: Artificial Intelligence, Machine Learning, Computer Vision, Form Recognizer, DAM.

ÍNDICE

INTRODUCCIÓN 1

1. Aspectos Generales 3

1.1. Descripción de la Empresa /Institución 3

1.1.1. Ubicación 3

1.1.2. Accesibilidad 3

1.1.3. Visión y misión de la empresa 4

1.1.4. Historia 5

2. Descripción De Una Actividad Específica 7

2.1. Computer Vision 8

2.1.1. Características Principales 9

2.1.2. Requisitos de entrada 10

2.2. Form Recognizer 10

2.2.1. Características Principales 12

2.2.2. Requisitos de entrada 12

2.3. Creación de Modelos Personalizados 14

2.3.1. Análisis de formatos de documentos 14

2.3.2. Creación del modelo 15

2.3.3. Etiquetado de pares clave-valor 17

2.3.4. Entrenamiento y pruebas del modelo 19

2.3.5. Integración 22

2.4. Modelado de base de datos 25

2.5. Diccionario de datos 27

3. ANTECEDENTES Y OBJETIVOS 38
3.1. Antecedentes internacionales 38
3.1.1. Simplificación de la incorporación de pacientes 38
3.1.2. Acelerando el tiempo de las conclusiones 39
3.1.3. Ahorrando tiempo en cada transacción 40
3.2. Antecedentes nacionales 41
3.2.1. Transformando recursos de manera sostenible 41
3.2.2. Automatización de procesos de facturación y contratos 42
3.3. Objetivos 43
3.3.1. Objetivo general 43
3.3.2. Objetivos específicos 44
4. Discusión y Aportes 45
5. Conclusiones y Recomendaciones 47
Bibliografía 48

ÍNDICE DE FIGURAS

Figura 1 Ubicación de CLI Gestiones Aduaneras S.A. 4

Figura 2 Proceso General de Lectura de Documentos 8

Figura 3 Ejemplo de extracción de texto OCR 9

Figura 4 Flujo de trabajo de Computer Vision 10

Figura 5 Factura Comercial analizada con Form Recognizer 11

Figura 6 Documentos en una Cuenta de almacenamiento de Azure 15

Figura 7 Formulario de creación de Modelos Personalizados 16

Figura 8 Espacio de trabajo de Form Recognizer Studio 17

Figura 9 Documento etiquetado listo para entrenar 18

Figura 10 Formulario para el entrenamiento de un modelo 19

Figura 11 Listado de modelos entrenado 20

Figura 12 Resultados del modelo entrenado 21

Figura 13 Resultados de pruebas con el modelo 22

Figura 14 Punto de conexión de Azure Form Recognizer 23

Figura 15 Resultado de análisis devuelto por API REST 24

Figura 16 Modelado de base de datos de Gestión Documentaria 26

Figura 17 Proceso detallado de carga de documentos 35

Figura 18 Ventana de carga de documentos 36

Figura 19 Documento clasificado automáticamente 36

Figura 20 Detalle extraído de una factura comercial 37

Figura 21 Funcionamiento de la aplicación MyHealthOne 39

Figura 22 Flujo de solución implementado por Standard Bank 41

Figura 23 Arquitectura Azure usada por EY..... 43

ÍNDICE DE TABLAS

Tabla 1 Instancias de Form Recognizer 13

Tabla 2 Diccionario de datos de archivo modelos_azure 27

Tabla 3 Diccionario de datos de archivo modelos_azure 28

Tabla 4 Diccionario de datos de archivo documentos_validaciones 29

Tabla 5 Diccionario de datos de archivo tipos_documento 30

Tabla 6 Diccionario de datos de archivo documentos_orden 31

Tabla 7 Diccionario de datos de archivo detalle_series 33

Tabla 8 Diccionario de datos de archivo usuarios 34

Tabla 9 Comparativo de servicios similares con Machine Learning 45

ACRÓNIMOS Y ABREVIATURAS

ERP	Planificación de Recursos Empresariales
DAM	Declaración Aduanero de Mercancías
SUNAT	Superintendencia Nacional de Aduanas y de Administración Tributaria
IA	Inteligencia Artificial
API	Interfaz de Programación de Aplicaciones
REST	Transferencia de Estado Representacional
JSON	Notación de Objeto de JavaScript

INTRODUCCIÓN

Un módulo de gestión documentaria dentro de un sistema con procesos administrativos es necesario en la actualidad. En este módulo se gestiona todos los documentos digitales y funciona como repositorio de documentos crucial dentro de una organización.

El proceso básico y general en este tipo de módulo debe ser de carga, clasificación y guardado del documento.

La empresa CLI GESTIONES ADUANERAS S.A. cuenta con un ERP construido sobre una plataforma web, en la cual se vienen desarrollando módulos según las necesidades que surjan dentro de la operatividad del día a día dentro de la organización, en los cuales uno de los módulos más importantes es el de "Gestión Documentaria". Este módulo permite la carga de documentos por cada despacho aperturado, los cuales posteriormente son usados por diferentes áreas como Administración, Transporte, Liquidación, Facturación entre otros.

Una de las primeras áreas que hace uso de los documentos cargados en este módulo es el área de liquidación, quienes se encargan de revisar y recopilar datos de distintos documentos para generar una DAM (Declaración Aduanera de Mercancías) que posteriormente se envía a SUNAT, por lo cual, el llenado de datos de esta declaración es parte fundamental del área de liquidación.

Dentro de este proceso, los liquidadores deben de abrir distintos documentos entre los cuales dos de lo más importantes son las Facturas Comerciales y los Documentos de Embarque. Dentro de las Facturas Comerciales se encuentran los productos que se deben de declarar, por lo cual, este documento puede tener una cantidad de hojas ilimitada. Así también, dentro de los Documentos de Embarque se encuentran detalles de contenedores y precintos que son necesarios para la DAM.

Ambos tipos de documentos pueden contar con una cantidad ilimitada de páginas, así también, estos documentos pueden ser digitales, escaneados o fotografías, por lo cual, los liquidadores no pueden simplemente copiar el texto del documento haciendo que el llenado de una DAM tome minutos, inclusive horas dependiendo la cantidad de productos y contenedores que tenga el despacho.

Otra problemática dentro del área es que, en muchas ocasiones, al trabajar con documentos escaneados, se tiene que digitalizar a mano cada documento, teniendo como consecuencia en muchas ocasiones errores de digitalización, por lo cual, al ser observadas por SUNAT conlleva a un proceso de rectificación de la DAM.

Por tales problemáticas, se implementó un lector de documentos que funciona sobre Machine Learning e Inteligencia Artificial dentro del módulo “Gestión Documentaria” que se encarga de extraer texto y datos específicos de distintos documentos que permite identificar qué tipo de documento es y almacenar los datos necesarios para que el área de liquidación pueda usarlo en la construcción de la DAM.

CAPÍTULO I

1. Aspectos Generales

1.1. Descripción de la Empresa /Institución

CLI GESTIONES ADUANERAS S.A. es una empresa que brinda servicios de logística de entrada y salida para carga general, productos perecederos y proyectos, con el apoyo de asociados de negocios locales y en el exterior.

1.1.1. Ubicación

- Departamento: PROV. CONST. DEL CALLAO
- Provincia: PROV. CONST. DEL CALLAO
- Distrito: CALLAO
- Dirección: AV. ELMER FAUCETT NRO. 2000 INT. D URB. INDUSTRIAL GRIMANESA

1.1.2. Accesibilidad

Para acceder a la ubicación exacta de la empresa, se puede tomar de referencia el Aeropuerto Internacional Jorge Chavez que se encuentra en la misma avenida Elmer Faucett. En dirección al sur, pasando el edificio Lima Cargo, se encuentra el paradero Naval. En este punto, cruzando la avenida Elmer Faucett se encuentran las instalaciones de la empresa CLI GESTIONES ADUANERAS S.A, según se muestra en la Figura 1.

Figura 1

Ubicación de CLI Gestiones Aduaneras S.A.



Nota. El punto rojo hace referencia a la ubicación exacta de CLI Gestiones Aduanera SA. Google Maps, 2022 (https://goo.gl/maps/SXbaiwBKoZuo3cpo8)

1.1.3. Visión y misión de la empresa

Visión

Ser la corporación más reconocida y competitiva en prestar servicios de gestión aduanera y logística integral, diferenciándonos por la calidad de nuestro servicio, alto profesionalismo de nuestros colaboradores e innovación tecnológica.

Misión

Fidelizar a nuestros socios comerciales y contribuir a que logren el éxito a partir de la prestación de un servicio óptimo, con alta tecnología, innovación y profesionalismo.

1.1.4. Historia

Corporación de Logística Integral (CLI) inicia sus actividades como agencia de aduanas (Géminis Agentes Afianzados de Aduanas) hace más de 30 años gracias a la estrategia de crecimiento vertical aplicada, el día de hoy cuenta con la capacidad de brindar servicios propios en toda la cadena logística, carga internacional, trámites aduaneros, transporte de carga local.

Su importante cartera de clientes corporativos, así como el hecho de haber figurado entre las 10 primeras agencias de aduanas desde 1995, es un claro esfuerzo por brindar siempre el mejor servicio, buscando constantemente la innovación y mejora continua como fuente de éxito.

Como producto de esta constante búsqueda, en 1995 fueron los primeros operadores logísticos en el país en brindar el servicio de importaciones y exportaciones bajo la modalidad puerta a puerta, específicamente al grupo telefónica. Esta experiencia les permitió aportar importantes mejoras en los tiempos de tránsito, con la consecuente reducción de costos fijos y variables de operación.

No contentos con ello, en el año 2001 fueron también los primeros en brindar la completa tercerización operativa del departamento de logística, manejando las operaciones puerta a puerta de productos terminados de Procter& Gamble en Perú, Ecuador y Colombia, con la permanente supervisión desde el exterior. Esta exitosa experiencia los ha llevado a establecer, igual que el caso anterior, menores tiempos de tránsito y grandes reducciones en costos.

Es importante indicar que, como consecuencia del éxito obtenido en el mercado local, a partir del año 2001, iniciaron el Plan de Desarrollo Regional, contando a la fecha con oficinas propias en Ecuador, las mismas que aplican la misma filosofía de servicio. (CLI, s.f.)

CAPÍTULO II

2. Descripción De Una Actividad Específica

La empresa, como lo redacta en su historia, se enfoca o da mucha importancia a la reducción de costos fijos y variables de la operación, por lo cual, para ello hace uso de la tecnología para la mejora de procesos constantemente.

Dentro de todos los proyectos tecnológicos implementados, se encuentra un proyecto ambicioso relacionado a la lectura y extracción de datos de documentos digitales, para su clasificación, facturación, declaración y entre otros procesos vitales de la empresa. Dicho proyecto denominado “Lectura de documentos con IA” será detallado en el presente informe.

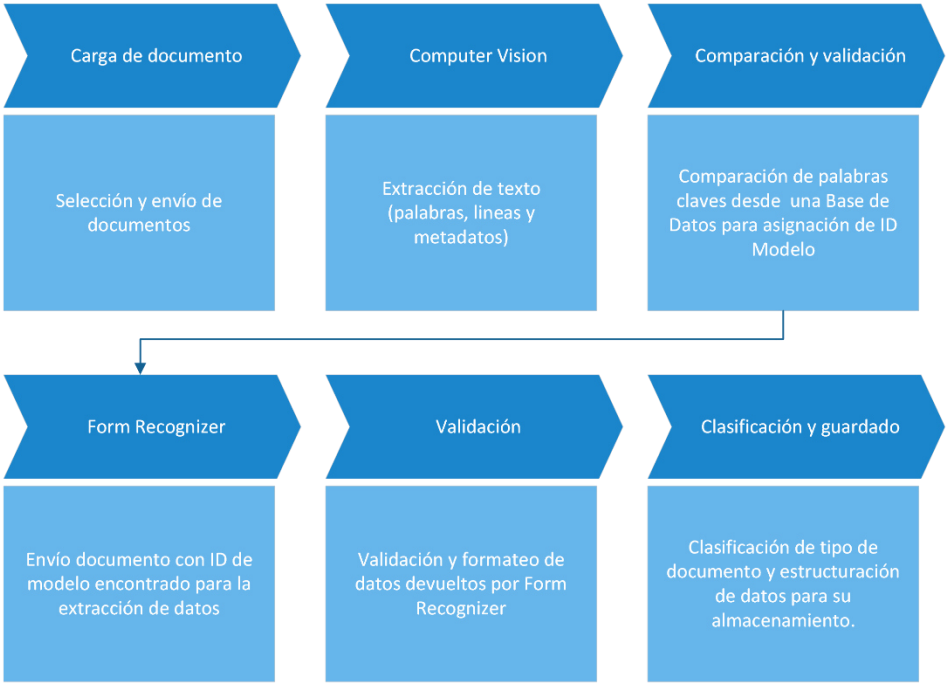
Para el desarrollo de dicho proyecto, se hizo uso de servicios en la nube como Computer Vision y Form Recognizer, que hace uso de un Aprendizaje Automatizado avanzado (Machine Learning), para crear modelos personalizados de documentos y hacer uso de ellos para los fines mencionados anteriormente. Ambos servicios pertenecen a la plataforma de Microsoft Azure.

Cabe resaltar que la empresa dentro de su ERP ya tenía implementado un módulo de gestión documentaria, por lo cual, este nuevo proceso de “Lectura de documentos con IA” se integró al dicho módulo “Gestión Documentaria”.

En la figura 2 se detallan los pasos del proceso implementado.

Figura 2

Proceso General de Lectura de Documentos



A continuación, se detallará la funcionalidad de los servicios Computer Vision y Form Recognizer para posteriormente explicar el uso de ellos dentro del módulo "Gestión Documentaria".

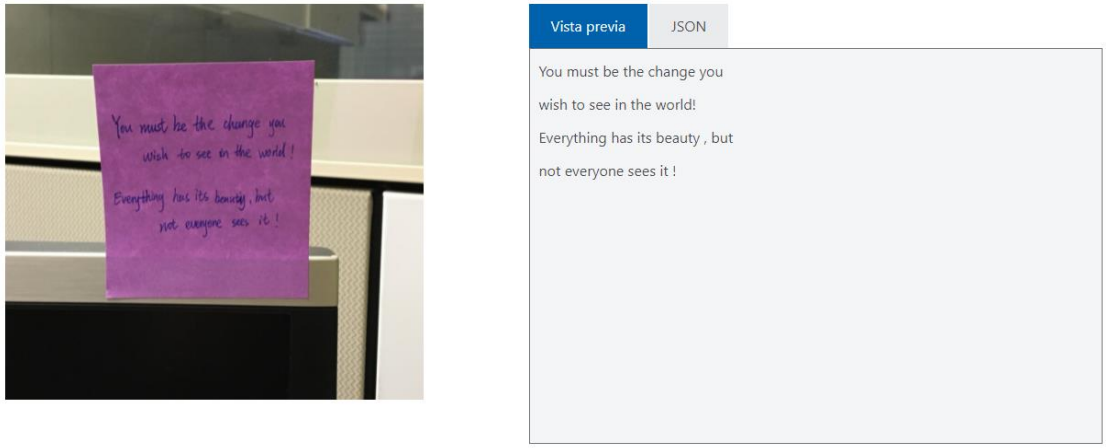
2.1. Computer Vision

Es un servicio de Azure que usa el procesamiento de datos visuales para etiquetar el contenido (desde objetos hasta conceptos), extraer texto impreso y escrito a mano, reconocer elementos conocidos como marcas y puntos de referencia y moderar el contenido. (Microsoft, s.f.)

En la figura 3 se muestra un ejemplo real del procesamiento de una imagen.

Figura 3

Ejemplo de extracción de texto OCR



Nota. Ejemplo usando el servicio de Azure Computer Vision, donde se muestra un resultado en texto plano y en formato JSON. Tomado de Microsoft Azure (https://azure.microsoft.com/es-es/products/cognitive-services/computer-vision/)

2.1.1. Características Principales

- Extracción de texto impreso en 122 idiomas.
- Extracción de texto escrita a mano en 7 idiomas.
- Líneas de texto y palabras con puntuaciones de ubicación y confianza.
- No se requiere identificación de idioma.
- Compatibilidad con idiomas mixtos, modo mixto (impresión y escritura a mano).
- Selección de páginas e intervalos de páginas de documentos grandes de varias páginas.
- Opción de orden de lectura natural para la salida de la línea de texto.
- Clasificación manuscrita para líneas de texto.

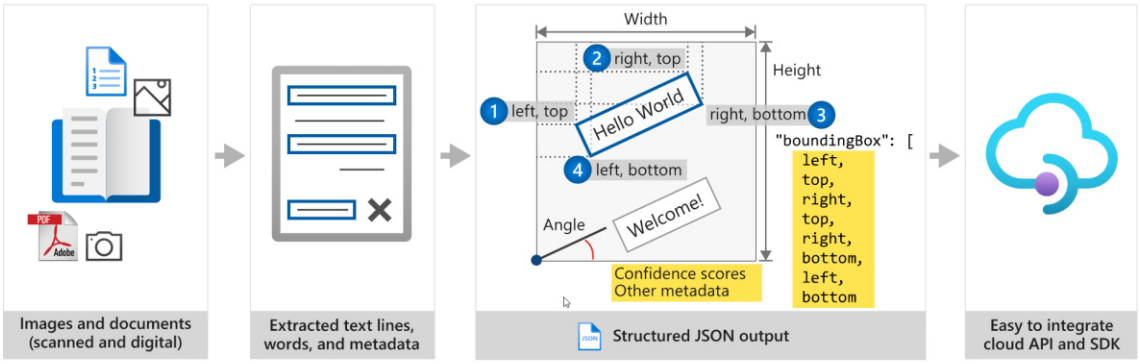
2.1.2. Requisitos de entrada

- Formatos de archivos admitidos: JPEG, PNG, BMP, PDF y TIFF.
- En el caso de los archivos PDF y TIFF, se procesan hasta 2000 páginas
- El tamaño de archivo debe ser inferior a 50 MB; y sus dimensiones, de al menos 50 x 50 píxeles y, como máximo, de 10 000 x 10 000 píxeles.

En la figura 4 se muestra el flujo de procesamiento de archivos de Computer Vision.

Figura 4

Flujo de trabajo de Computer Vision



Nota. Flujo de trabajo usando Computer Vision. Extraído de la documentación oficial de Form Recognizer (https://learn.microsoft.com/es-es/azure/cognitive-services/computer-vision/overview-ocr)

2.2. Form Recognizer

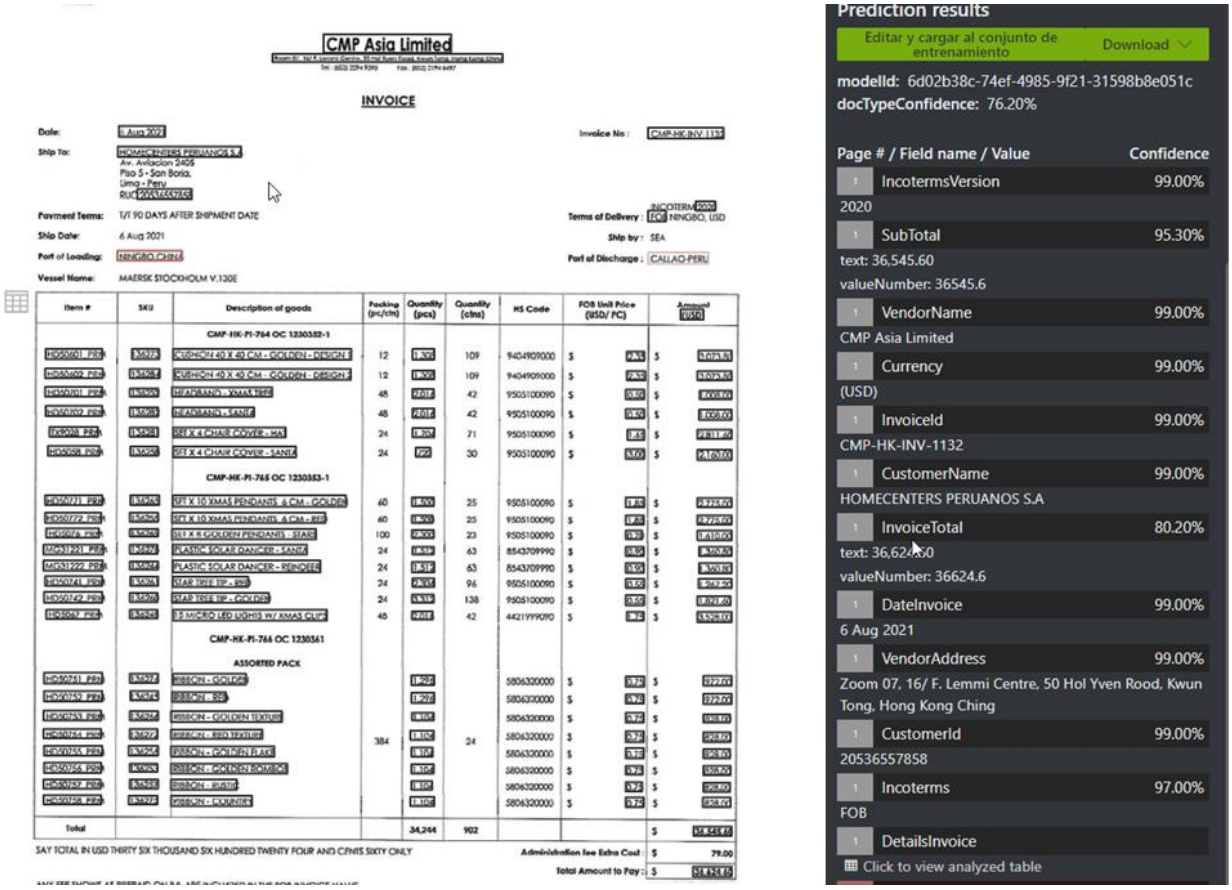
Form Recognizer es otro servicio de Azure que aplica aprendizaje automático avanzado para la extracción de texto con precisión, pares clave-valor, tablas y estructuras de los documentos.

Con solo algunos ejemplos, se puede adaptar Form Recognizer para que reconozca documentos, tanto en el entorno local como en la nube.

En la figura 5 se muestra el resultado de una factura comercial analizado por Form Recognizer.

Figura 5

Factura Comercial analizada con Form Recognizer



Nota. A la izquierda si visualiza el documento analizado, en donde las zonas con rectángulos oscuros hacen referencia a los valores encontrados correspondiente a cada etiqueta. A

la derecha se muestran las etiquetas del modelo con sus respectivos valores y porcentajes de confianza.

2.2.1. Características Principales

Form Recognizer identifica, extrae y analiza fácilmente los siguientes datos de los documentos:

- Estructura y contenido de la tabla.
- Elementos de formulario y valores de campo.
- Texto alfanumérico escrito y escrito a mano.
- Relaciones entre elementos.
- Pares clave-valor.
- Ubicación de elementos con coordenadas de rectángulo de selección.

2.2.2. Requisitos de entrada

- Para obtener unos resultados óptimos, se debe proporcionar una foto clara o una digitalización de alta calidad por documento.
- Formatos de archivo admitidos: JPEG, PNG, BMP, TIFF y PDF (texto insertado o digitalizado). Los PDF insertados de texto son mejores para eliminar la posibilidad de error en la extracción de caracteres y en la ubicación.
- En el caso de PDF y TIFF, se pueden procesar hasta 2000 páginas.
- El tamaño del archivo debe ser inferior a 50 MB.
- Las imágenes deben tener unas dimensiones entre 50 x 50 píxeles y 10 000 x 10 000 píxeles.

- Los archivos PDF tienen unas dimensiones de hasta 17 x 17 pulgadas, lo que corresponde a los tamaños de papel Legal o A3, o más pequeños.
- El tamaño total de los datos de entrenamiento es de 500 páginas o menos.
- Si los archivos PDF están bloqueados con contraseña, debe desbloquearlos antes de enviarlos.

Como se muestra en la tabla 1, este servicio cuenta con 6 instancias de uso, por lo cual, según el análisis realizado, el que más de se adaptó a la necesidad del negocio fue la instancia de "Modelo personalizado".

Tabla 1

Instancias de Form Recognizer

Instancia	Características
API de diseño	Extracción y análisis de texto, marcas de selección y estructuras de tabla, junto con sus coordenadas de rectángulo de selección, de formularios y documentos.
Modelo personalizado	Extracción y análisis de datos de formularios y documentos específicos de distintos datos empresariales y casos de uso.
Modelo de factura	Procesamiento automatizado de datos y extracción de información clave de facturas.
Modelo de recibo	Procesamiento automatizado de datos y extracción de información clave de recibos de compra.
Modelo de documentación de id.	Procesamiento automatizado de datos y extracción de información clave de permisos de conducir de EE. UU. y pasaportes internacionales.
Modelo de tarjeta de presentación	Procesamiento automatizado de datos y extracción de información clave de tarjetas de presentación.

Teniendo conocimiento de los servicios de Azure usados en este proceso, se procederá a detallar paso a paso el análisis, creación y entrenamiento de modelos personalizados llevados a cabo.

2.3. Creación de Modelos Personalizados

2.3.1. Análisis de formatos de documentos

Como primer paso, se necesitó analizar los documentos que serían usados para la creación y entrenamiento de los modelos. Según la documentación de Form Recognizer, se recomienda seleccionar formatos de documentos que sean muy similares o iguales para la creación de un modelo personalizado. (Microsoft, s.f.)

Para el desarrollo del presente informe, se usará como ejemplo el documento de Factura Comercial. Los formatos de estos documentos son distintos según el proveedor, por lo cual, se seleccionaron por lo menos 5 facturas de cada proveedor para la creación de los modelos. Si bien, estas facturas tenían similitud en el formato y estructura de sus datos, se eligieron facturas que fueron lo más distinto posible en el contenido, es decir, facturas que tenían detalles de 1, 5, 10 o más líneas, haciendo que la cantidad de páginas de los documentos sean distintas. Esto permitió al modelo aprender los diferentes casos que pueda presentar una factura por proveedor.

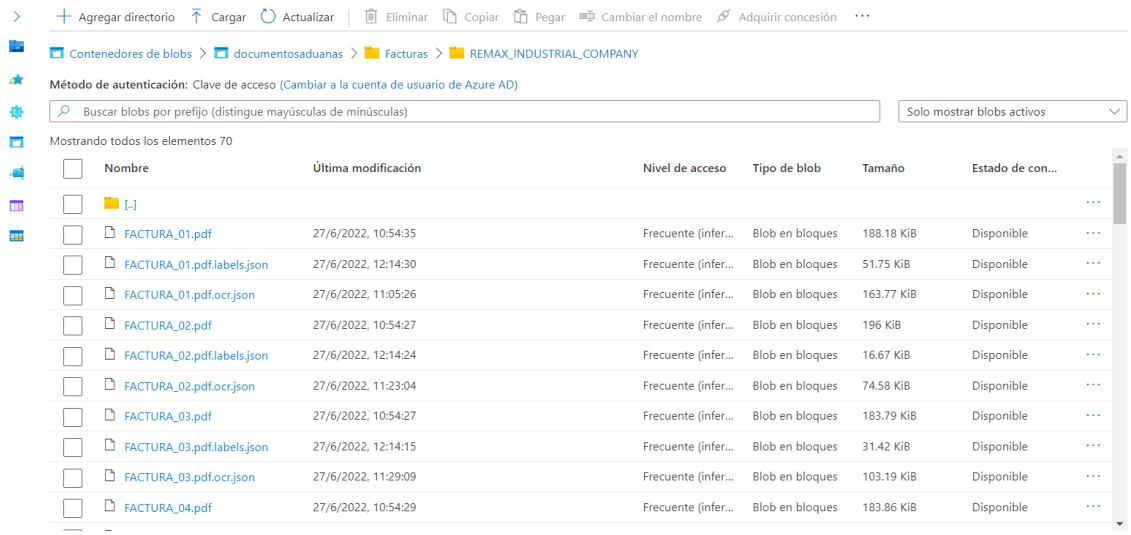
Cabe tener en cuenta que según la documentación de Form Recognizer, para crear un modelo se necesitan como mínimo 5 documentos de ejemplo que tengan el mismo formato y estructura. (Microsoft, s.f.)

Una vez creado y entrenado el modelo, según la precisión del resultado, se debe de agregar más documentos de ejemplo hasta que la precisión de los resultados sea de por lo menos el 99%.

Una vez seleccionado los 5 modelos de factura, se procedió a subirlos a un contenedor del servicio Cuenta de almacenamiento de Azure. Dentro del contenedor, se creó una carpeta de facturas y dentro de ella una carpeta por cada proveedor como se visualiza en la figura 6.

Figura 6

Documentos en una Cuenta de almacenamiento de Azure



El contenedor fue configurado con los permisos necesarios para que Form Recognizer tuviera acceso a los documentos almacenados.

2.3.2. Creación del modelo

Para la creación de un nuevo modelo, se hizo uso de la herramienta Form Recognizer Studio desarrollada por el mismo equipo de Azure para facilitar el uso del servicio de Form Recognizer.

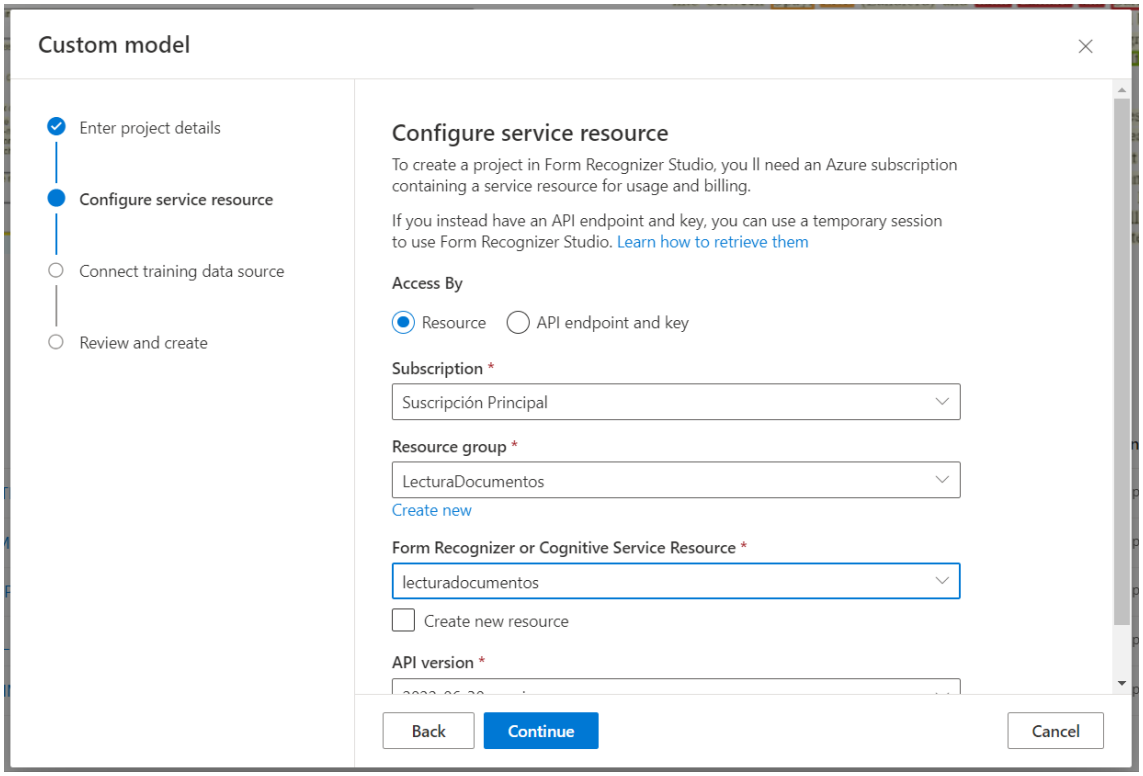
Dentro de la plataforma, se procedió a crear un nuevo modelo personalizado donde se ingresaron algunos datos requeridos:

- Nombre del modelo
- Versión de API
- Ruta del directorio en el contenedor de Almacenamiento
- Otros

En la figura 7 se muestra el formulario de un nuevo modelo personalizado.

Figura 7

Formulario de creación de Modelos Personalizados



Una vez ingresado los datos requeridos y creado el modelo, la herramienta muestra el espacio de trabajo para crear etiquetas y asignarles sus valores correspondientes.

En la figura 8 se muestra el espacio de trabajo donde se visualizan etiquetas de color amarillo indicado que es texto analizado con OCR y que puede ser usado para el etiquetado.

Figura 8

Espacio de trabajo de Form Recognizer Studio



Nota. A la izquierda se muestran los documentos de muestra cargados en la Cuenta de Almacenamiento, en el centro el documento con los textos analizados con OCR y la derecha el apartado de etiquetas.

2.3.3. Etiquetado de pares clave-valor

Como lo mencionado anteriormente, la lógica que aplica el aprendizaje automatizado de Form Recognizer se basa en pares de clave y valor. Es decir, se necesita crear etiquetas claves como, por ejemplo: Nombre del proveedor, Nro. Factura, Fecha de Emisión, entre otros. En cada etiqueta creada se puede especificar el tipo de dato que puede ser, por ejemplo: Fecha, Numero,

moneda, texto, etc. Así también, la herramienta permite crear tablas dinámicas, las cuales son útiles para el detalle de la factura, ya que la cantidad de series o líneas son variables. En esta tabla se debe de indicar el encabezado que puede tener campos como: Nro item, código, descripción, precio unitario, precio total, etc.

Una vez creada las etiquetas, se procedió a asignarle su valor a cada una. Mediante Form Recognizer Studio este trabajo es muy simple, ya que se sólo se deben seleccionar las palabras en el documento y con un clic indicar a que etiqueta pertenece. El resultado se muestra en la figura 9.

Figura 9

Documento etiquetado listo para entrenar



Nota. Documento etiquetado. Deben etiquetarse todos los documentos para proceder con el entrenamiento.

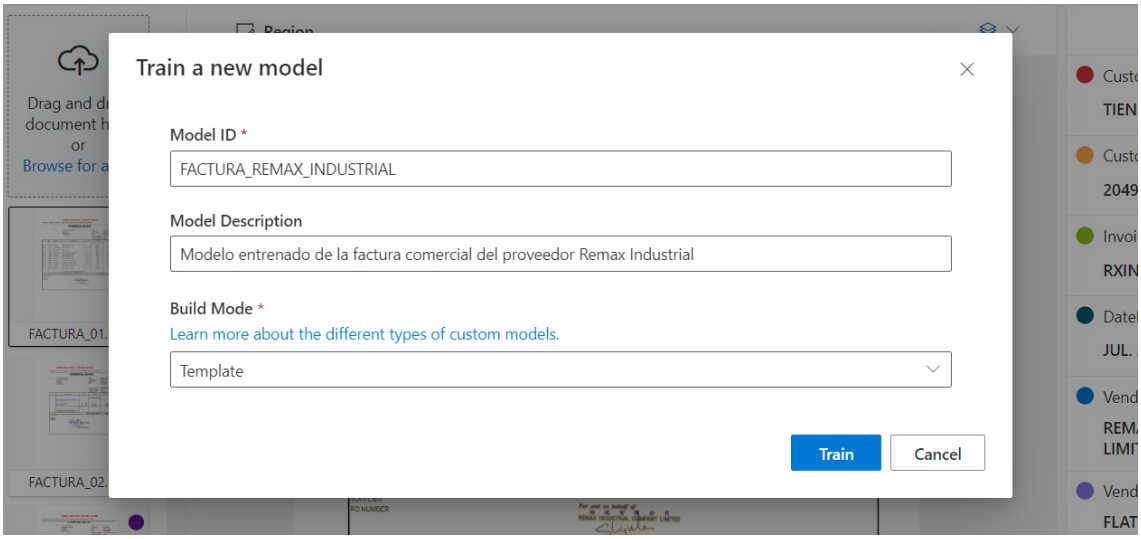
2.3.4. Entrenamiento y pruebas del modelo

El proceso de etiquetado de pares clave-valor se debe realizar con todos los documentos cargados. Una vez terminado con el etiquetado de todos los documentos, se procedió a entrenar el modelo.

Para ello, Form Recognizer solicita un ID Modelo, que se usará como un identificador único, una descripción y el tipo de entrenamiento. En este caso se hizo uso del tipo de entrenamiento por Template como se muestra en la figura 10.

Figura 10

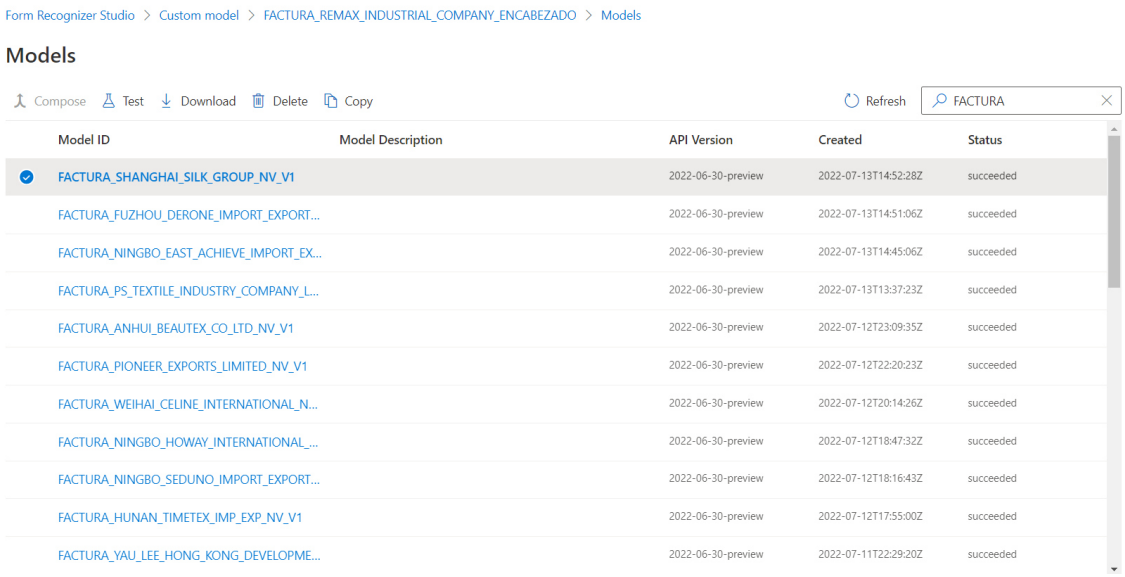
Formulario para el entrenamiento de un modelo



El entrenamiento suele demorar algunos minutos y se ejecuta en segundo plano. En la figura 11 se muestran los modelos entrenados.

Figura 11

Listado de modelos entrenados

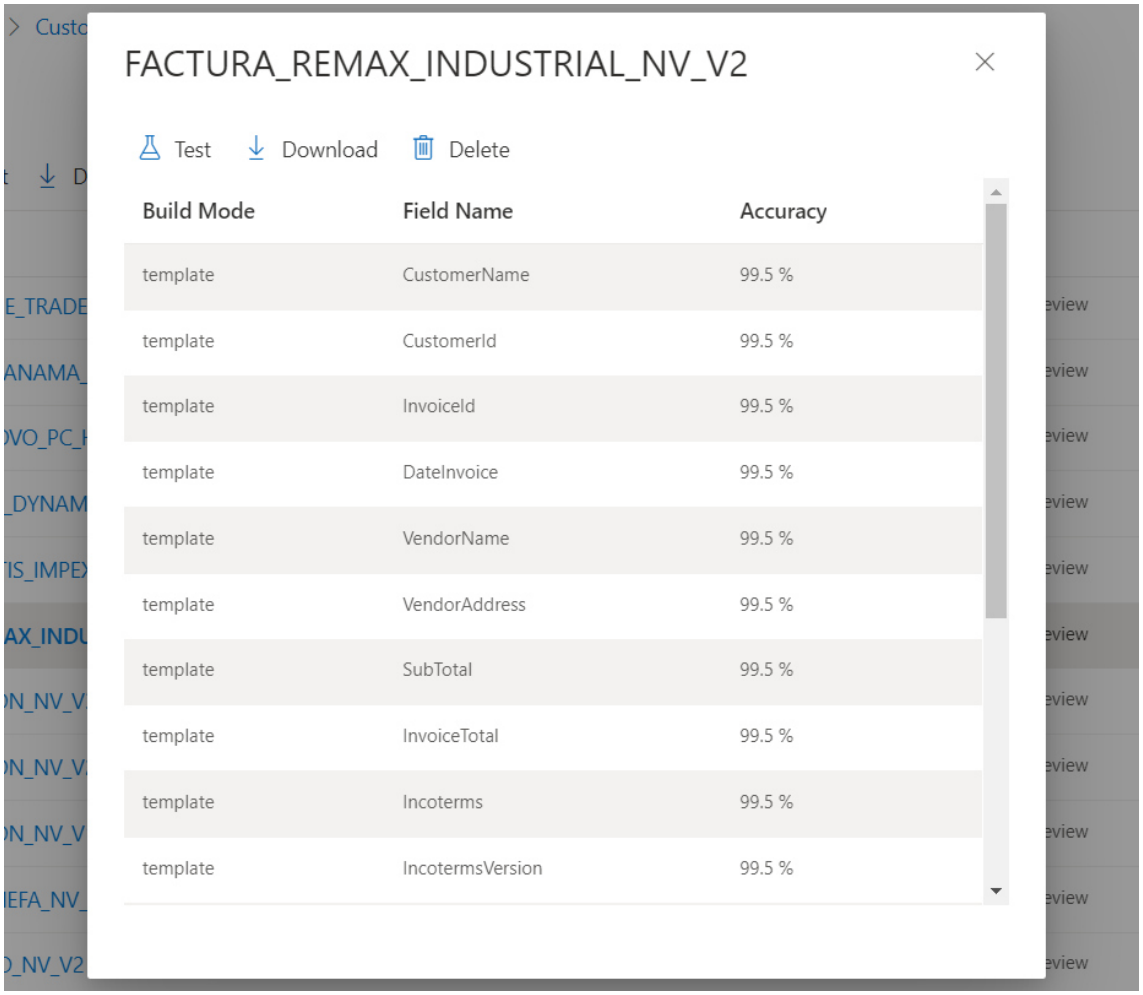


Nota. En el listado figura el estado del modelo (running, succeeded o error), la versión de API que se usó y la fecha y hora de creación.

Dentro de cada modelo se visualiza el detalle de los resultados de entrenamiento. En el detalle se debe de analizar los porcentajes de precisión por cada etiqueta y reforzar las etiquetas que presentan una precisión baja. Dentro de la ventana también se encuentra la opción de “Test” para realizar pruebas con el modelo entrenado como se muestra en la figura 12.

Figura 12

Resultados del modelo entrenado

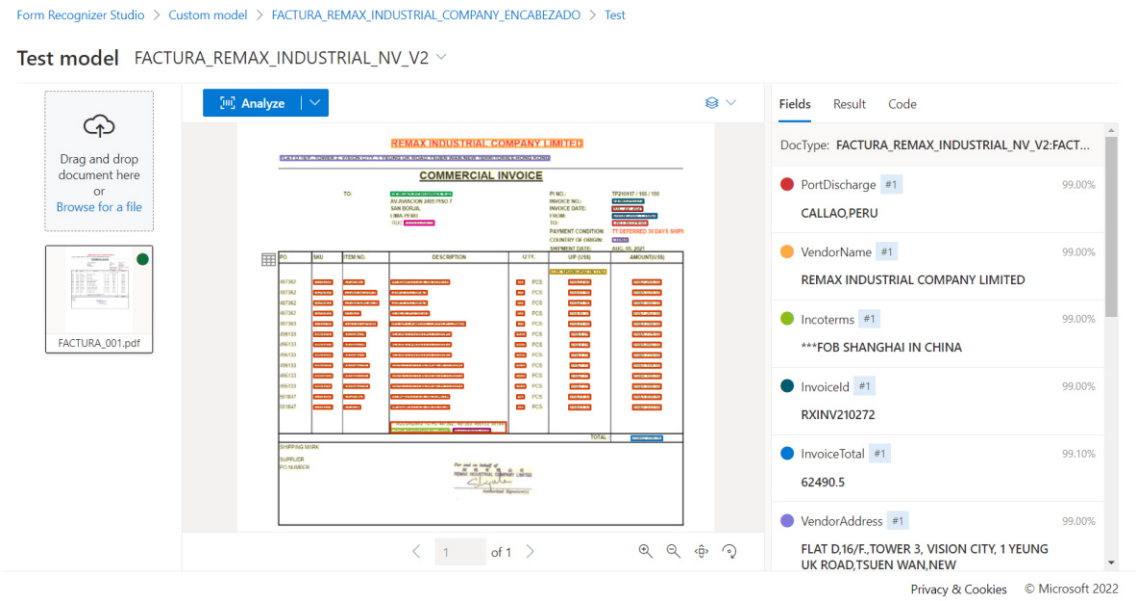


Nota. Los resultados del entrenamiento muestran los porcentajes de precisión por cada etiqueta.

Los resultados de las pruebas realizadas con el modelo se visualizan de la figura 13.

Figura 13

Resultados de pruebas con el modelo



Nota. Al lado derecho se muestran los resultados de análisis del documento.

En esta fase se realizó la mayor cantidad de pruebas con distintas facturas. En caso de que una factura no alcance un grado de confianza de 99%, se debe de cargarla en la Cuenta de Almacenamiento, etiquetarla y entrenar el modelo nuevamente.

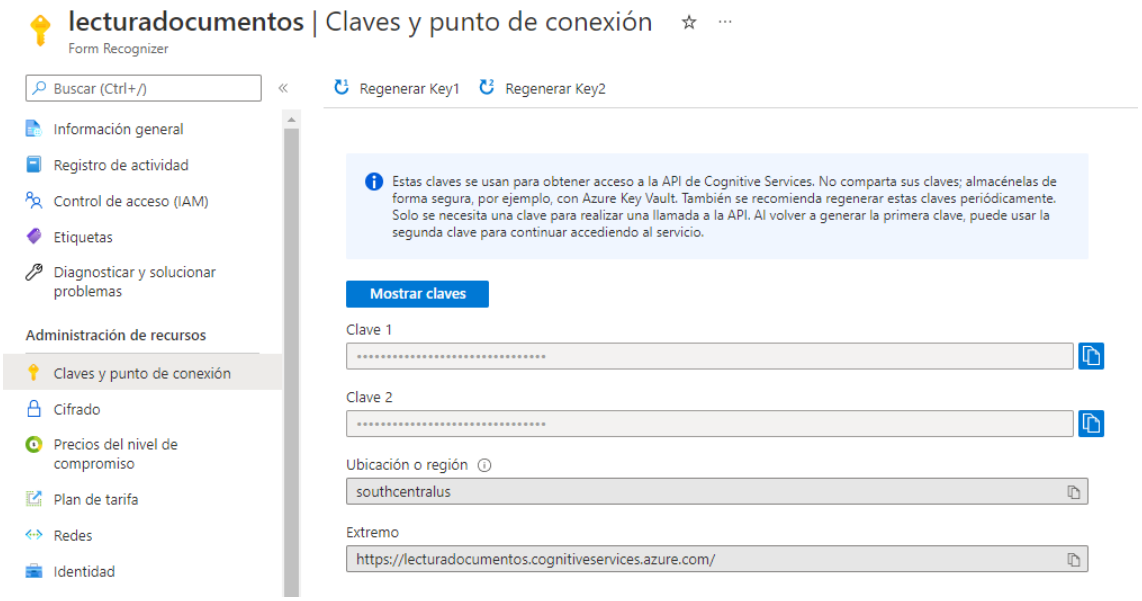
2.3.5. Integración

Para hacer uso del modelo entrenado desde el módulo de "Gestión Documentaria" se hizo uso de la API REST brindada por Azure Form Recognizer.

Para ello se realizó una petición tipo POST al endpoint del servicio usando la llave privada que nos brinda Azure Form Recognizer como se muestra en la figura 14.

Figura 14

Punto de conexión de Azure Form Recognizer

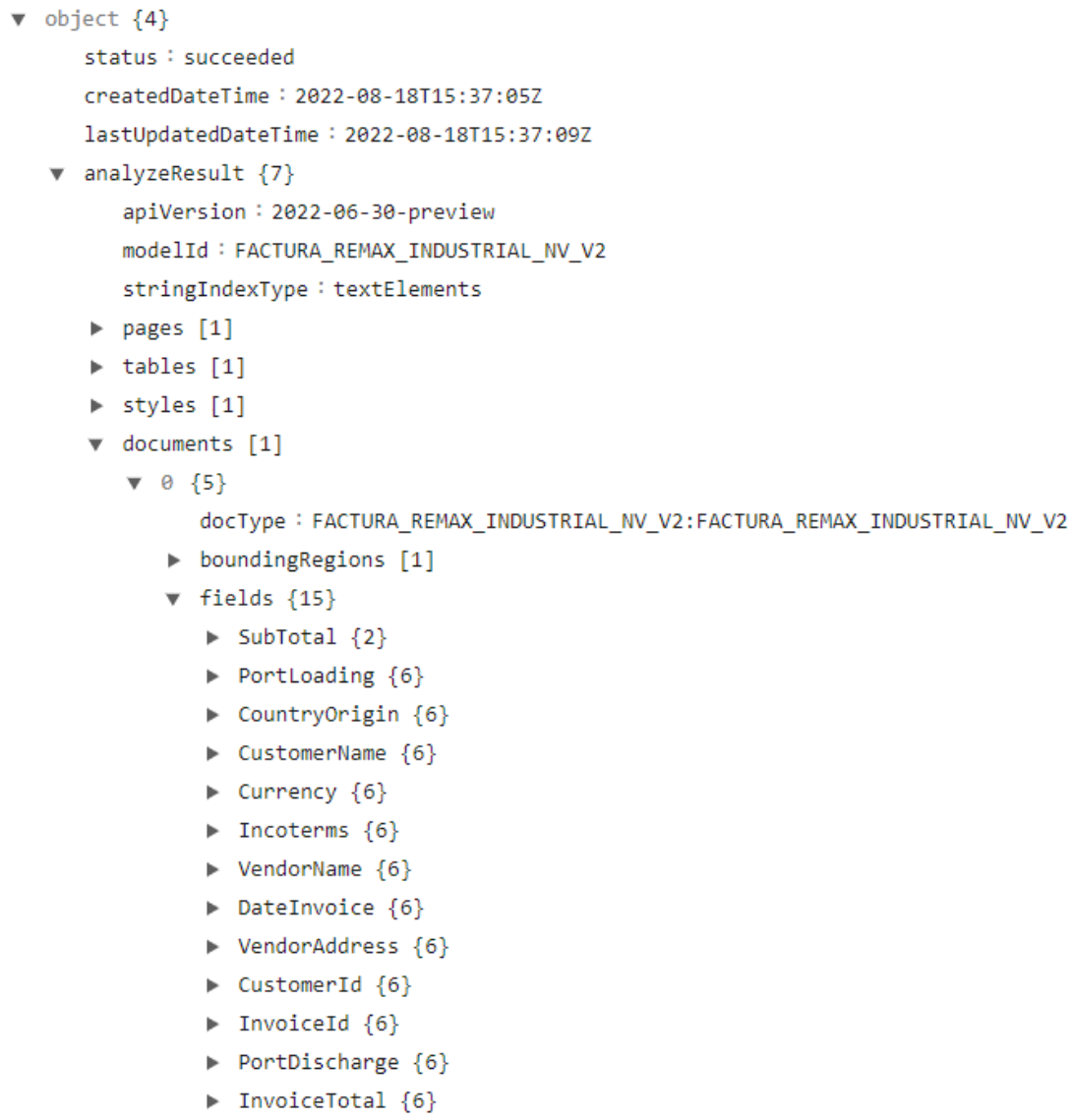


Junto a la petición se envió el ID de modelo creado, el archivo a analizar y la versión de la API que se usará para el análisis. Esta petición devuelve como respuesta un ID de Resultado que se envía en otra petición GET para obtener los resultados del análisis del documento.

Si todo marcha de la manera correcta, Form Recognizer devuelve un JSON con todos los datos obtenidos del análisis como se muestra en la figura 15.

Figura 15

Resultado de análisis devuelto por API REST



De esa manera se consume el servicio para analizar los documentos requeridos.

Como se indicó anteriormente, este proceso fue integrado al módulo “Gestión Documentaria”, por lo cual, el análisis de los documentos se realiza cuando el usuario carga un documento en dicho módulo.

En el momento que el usuario carga el documento, este debería analizarse para extraer sus datos y clasificarlo.

En este punto surgió un problema ya que el usuario puede cargar cualquier documento PDF, pero, como la API de Form Recognizer requiere de un ID modelo para realizar el análisis, se necesitaba primero identificar si el documento pertenecía a una factura comercial para poder enviar el ID de modelo correspondiente a Azure.

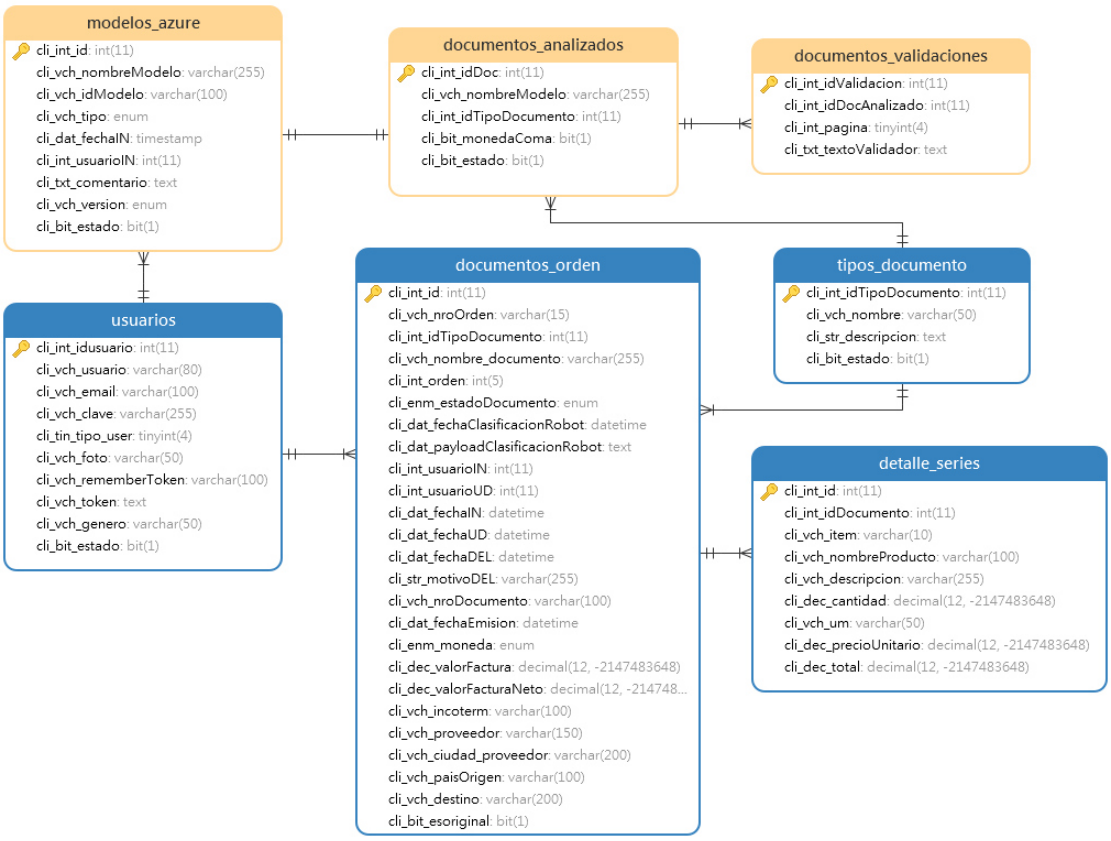
Para ello, se hizo uso del servicio Computer Visión, el cual realiza el trabajo de extraer todas las palabras y líneas del documento las cuales se pueden analizar y validar mediante palabras claves de una tabla en una base de datos, es decir, se implementó una tabla que contienen palabras únicas de un documento con un formato único, las cuales se usan para saber si el documento cargado contiene dichas palabras. En caso de que la cantidad de palabras coincidan, se sabrá que documento es y que ID de modelo usar.

2.4. Modelado de base de datos

En la figura 16, se muestra el modelado de base de datos implementado en el módulo “Gestión Documentaria”.

Figura 16

Modelado de base de datos de Gestión Documentaria



Nota. Las tablas de color naranja pertenecen al proceso de análisis con los servicios de Azure. Las tablas de color azul pertenecen al proceso del módulo Gestión Documentaria.

2.5. Diccionario de datos

Tabla 2

Diccionario de datos de archivo modelos_azure

Nombre de Archivo	Descripción		
modelos_azure	Tabla donde se registran los modelos creados en Azure Form Recognizer Studio		
Campo	Tipo	Tamaño	Descripción
cli_int_id	Numérico	11	Clave autogenerado del modelo
cli_vch_nombreModelo	Texto	255	Nombre del modelo
cli_vch_idModelo	Texto	100	Identificador del modelo Azure
cli_vch_tipo	Texto	50	Tipo del modelo Azure
cli_dat_fechaIN	Fecha	20	Fecha y hora de registro
cli_int_usuarioIN	Numérico	11	Clave del usuario
cli_txt_comentario	Texto	300	Comentario del modelo
cli_vch_version	Texto	5	Versión de la API Azure usado
cli_bit_estado	Numérico	1	Estado del modelo (Activo o inactivo)
Relaciones	Campos Clave		
usuarios.cli_int_idusuario con cli_int_usuarioIN	cli_int_id, cli_vch_nombreModelo		

Tabla 3

Diccionario de datos de archivo documentos_analizados

Nombre de Archivo	Descripción		
documentos_analizados	Tabla donde se registran los documentos que se analizarán con Form Recognizer cuando se cargue un archivo en el módulo de GESTION DOCUMENTARIA.		
Campo	Tipo	Tamaño	Descripción
cli_int_idDoc	Numérico	11	Clave autogenerado del documento analizado
cli_vch_nombreModelo	Texto	255	Nombre del modelo
cli_int_idTipoDocumento	Numérico	11	Identificador del tipo de documento
cli_bit_monedaComa	Numérico	1	Indicador de coma decimal en moneda del documento
cli_bit_estado	Numérico	1	Estado del documento analizado (Activo o inactivo)
Relaciones	Campos Clave		
modelos_azure.cli_vch_nombreModelo con cli_vch_nombreModelo tipos_documento.cli_int_idTipoDocumento con cli_int_idTipoDocumento	cli_int_idDoc, cli_vch_nombreModelo, cli_int_idTipoDocumento		

Tabla 4

Diccionario de datos de archivo documentos_validaciones

Nombre de Archivo	Descripción		
documentos_validaciones	Tabla donde se registran las palabras claves a buscar en un documento para posteriormente enviarlo a Form Recognizer. Tabla usada con el servicio Computer Vision.		
Campo	Tipo	Tamaño	Descripción
cli_int_idValidacion	Numérico	11	Clave autogenerado de la validación
cli_int_idDocAnalizado	Numérico	11	Identificador del documento analizado
cli_int_pagina	Numérico	4	Número de página donde buscar
cli_txt_textoValidador	Texto	100	Palabra o texto por buscar en el documento
Relaciones	Campos Clave		
documentos_analizados.cli_int_idDoc con cli_int_idDocAnalizado	cli_int_idValidacion, cli_int_idDocAnalizado		

Tabla 5

Diccionario de datos de archivo tipos_documento

Nombre de Archivo	Descripción		
tipos_documento	Tabla donde se registran los tipos de documentos para el módulo de GESTIÓN DOCUMENTARIA		
Campo	Tipo	Tamaño	Descripción
cli_int_idTipoDocumento	Numérico	11	Clave autogenerado del tipo de documento
cli_vch_nombre	Texto	50	Nombre del tipo de documento
cli_str_descripcion	Texto	300	Descripción del tipo de documento
cli_bit_estado	Numérico	1	Estado del tipo de documento (Activo o inactivo)
Relaciones	Campos Clave		
	cli_int_idTipoDocumento		

Tabla 6

Diccionario de datos de archivo documentos_orden

Nombre de Archivo	Descripción		
documentos_orden	Tabla donde se registran los documentos cargados en el módulo GESTIÓN DOCUMENTARIA		
Campo	Tipo	Tamaño	Descripción
cli_int_id	Numérico	11	Clave autogenerado del documento
cli_vch_nroOrden	Texto	15	Número de orden aduanero interno
cli_int_idTipoDocumento	Numérico	11	Identificador del tipo de documento
cli_vch_nombre_documento	Texto	255	Nombre del documento
cli_int_orden	Numérico	5	Orden del documento
cli_enm_estadoDocumento	Texto	10	Estado de análisis del documento
cli_dat_fechaClasificacionRobot	Fecha	20	Fecha en que se analizó el documento
cli_dat_payloadClasificacionRobot	Texto	2000	Json generado por el análisis del documento
cli_int_usuarioIN	Numérico	11	Identificador del usuario quien registra el documento
cli_int_usuarioUD	Numérico	11	Identificador del usuario quien actualiza el documento
cli_dat_fechaIN	Fecha	20	Fecha en que se registra el documento
cli_dat_fechaUD	Fecha	20	Fecha en que se actualiza el documento
cli_dat_fechaDEL	Fecha	20	Fecha en que se elimina el documento

cli_str_motivoDEL	Texto	500	Motivo de eliminación del documento
cli_vch_nroDocumento	Texto	100	Número del documento
cli_dat_fechaEmision	Fecha	20	Fecha de emisión del documento
cli_enm_moneda	Texto	10	Moneda de emisión del documento
cli_dec_valorFactura	Numérico	12	Valor da la factura
cli_dec_valorFacturaNeto	Numérico	12	Valor neto de la factura
cli_vch_incoterm	Texto	100	Incoterms del documento
cli_vch_proveedor	Texto	150	Proveedor de la factura
cli_vch_ciudad_proveedor	Texto	200	Ciudad del proveedor de la factura
cli_vch_paisOrigen	Texto	100	País de origen de la factura
cli_vch_destino	Texto	200	País destino de la factura
cli_bit_esoriginal	Numérico	1	Indica si es un documento original o borrador

Relaciones	Campos Clave
tipos_documento.cli_int_idTipoDocumento con cli_int_idTipoDocumento usuarios.cli_int_idusuario con cli_int_usuarioIN, cli_int_usuarioUD	cli_int_id, cli_int_idTipoDocumento, cli_int_usuarioIN, cli_int_usuarioUD

Tabla 7

Diccionario de datos de archivo detalle_series

Nombre de Archivo	Descripción		
detalle_series	Tabla donde se registra el detalle de las facturas en caso el documento sea de este tipo de documento		
Campo	Tipo	Tamaño	Descripción
cli_int_id	Numérico	11	Clave autogenerado del modelo
cli_int_idDocumento	Numérico	11	Identificador del documento
cli_vch_item	Texto	10	Número de ítem del producto
cli_vch_nombreProducto	Texto	100	Nombre del producto
cli_vch_descripcion	Texto	255	Descripción del producto
cli_dec_cantidad	Numérico	12	Cantidad del producto
cli_vch_um	Texto	50	Unidad de medida
cli_dec_precioUnitario	Numérico	12	Precio unitario del producto
cli_dec_total	Numérico	12	Precio total del producto
Relaciones	Campos Clave		
documentos_orden.cli_int_idDocumento con cli_int_idDocumento	cli_int_id, cli_int_idDocumento		

Tabla 8

Diccionario de datos de archivo usuarios

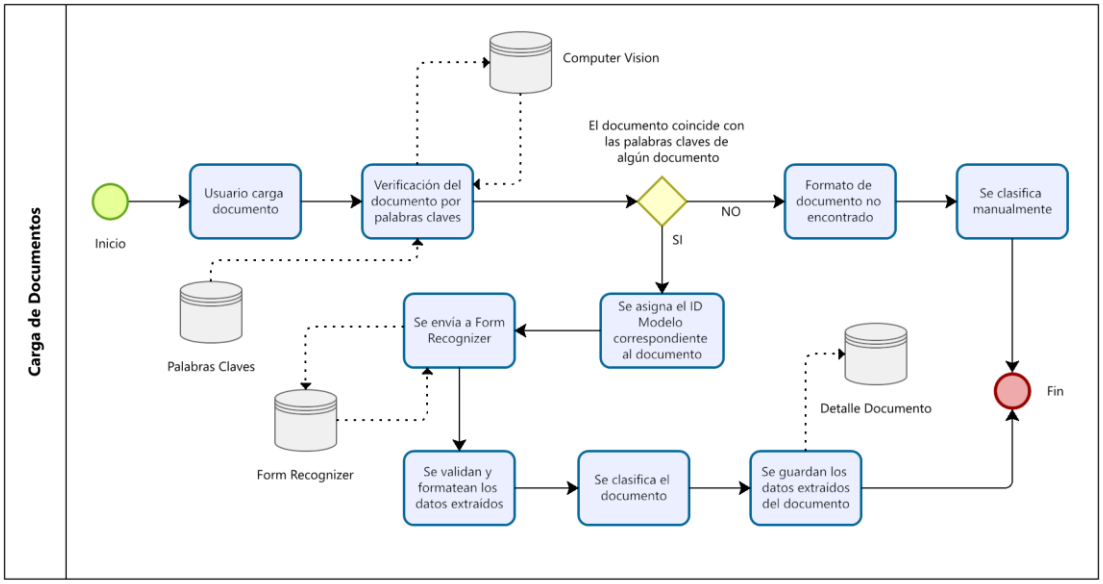
Nombre de Archivo	Descripción		
usuarios	Tabla donde se registran los usuarios del sistema		
Campo	Tipo	Tamaño	Descripción
cli_int_idusuario	Numérico	11	Clave autogenerado del modelo
cli_vch_usuario	Texto	80	Nombre de usuario
cli_vch_email	Texto	100	Correo electrónico del usuario
cli_vch_clave	Texto	255	Clave del usuario
cli_tin_tipo_user	Numérico	4	Tipo de usuario
cli_vch_foto	Texto	50	Nombre de la foto del usuario
cli_vch_rememberToken	Texto	100	Token de olvido de contraseña
cli_vch_token	Texto	200	Token del usuario
cli_vch_genero	Texto	50	Genero del usuario
cli_bit_estado	Numérico	1	Estado del usuario (Activo o inactivo)
Relaciones		Campos Clave	
		cli_int_idusuario	

Con esta implementación, el módulo “Gestión Documentaria” ya se encuentra listo para analizar los documentos requeridos.

En la figura 17 se muestra a detalle el proceso de carga de un documento.

Figura 17

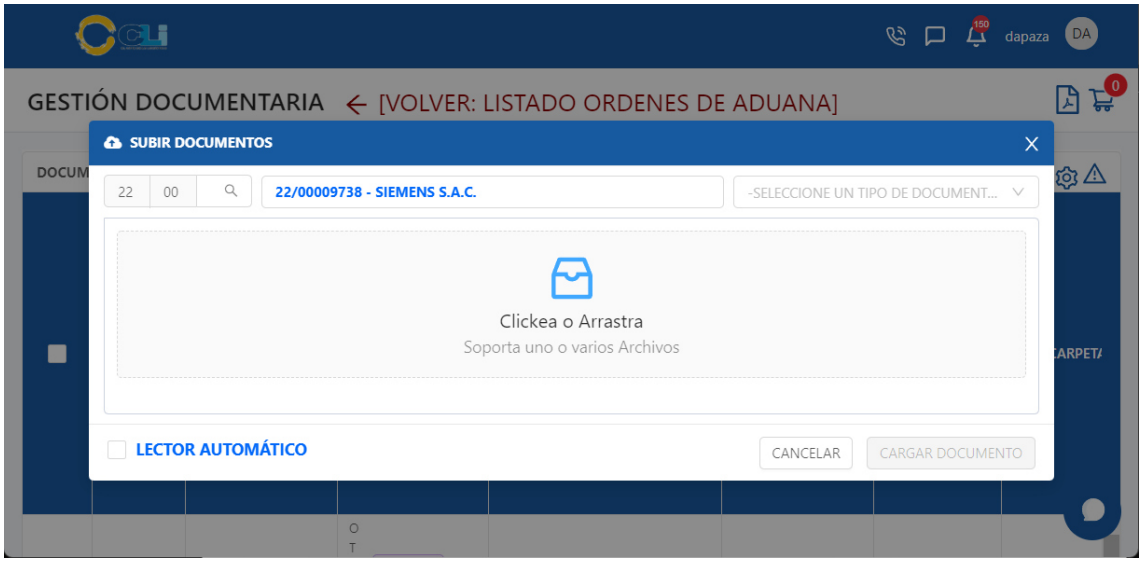
Proceso detallado de carga de documentos



La figura 18 corresponde a la ventana de carga de documentos.

Figura 18

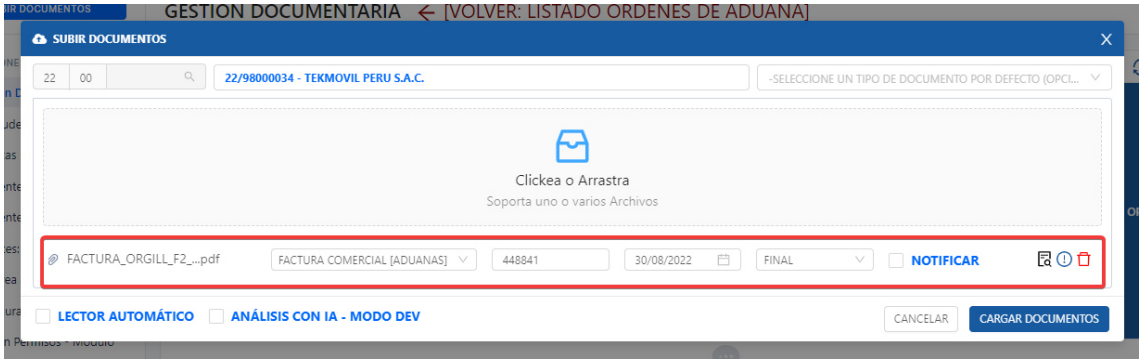
Ventana de carga de documentos



Si el documento cargado pertenece a algún formato trabajado con Form Recognizer, se realiza la clasificación del tipo de documento, en este caso Factura Comercial, como se muestra en la figura 19.

Figura 19

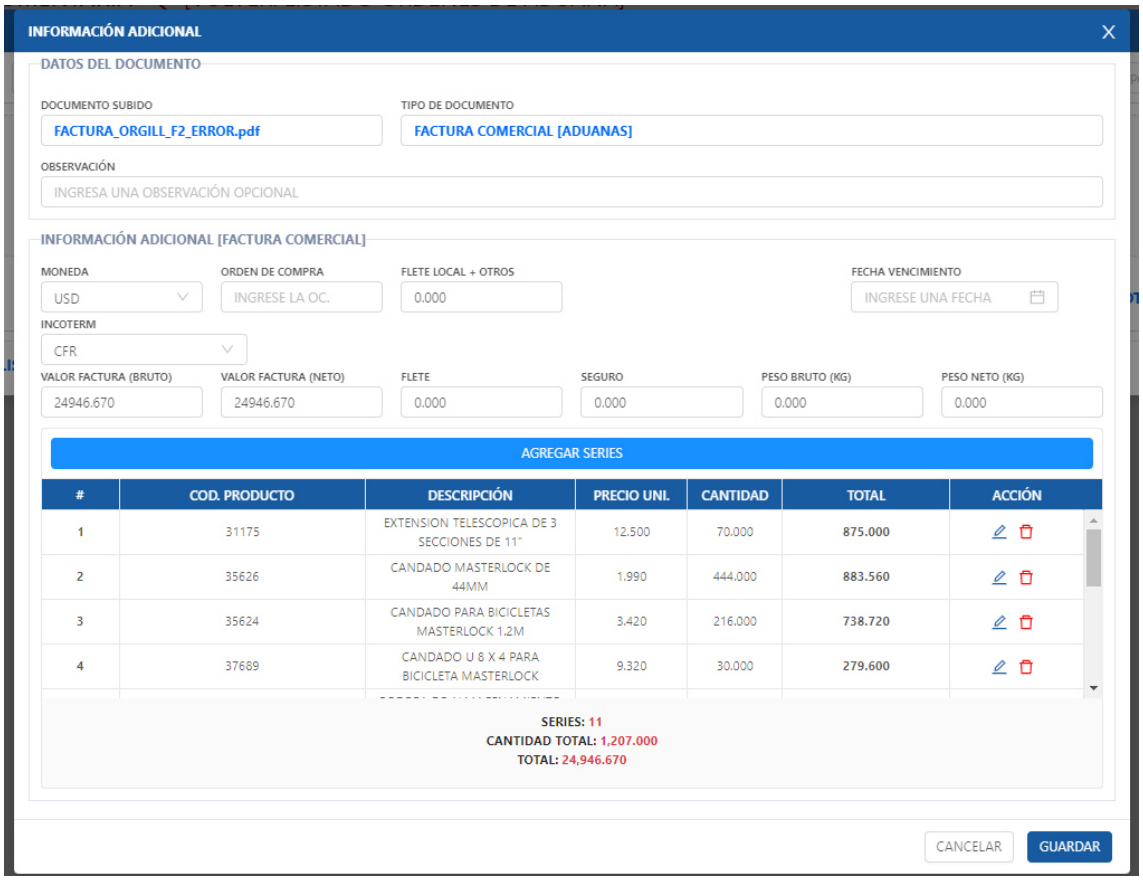
Documento clasificado automáticamente



Así también, en la figura 20 se visualiza el detalle de la factura. Todos estos datos se guardan en una base de datos que posteriormente será usada para que el área de liquidaciones declare el despacho aduanero a SUNAT.

Figura 20

Detalle extraído de una factura comercial



CAPITULO III

3. ANTECEDENTES Y OBJETIVOS

3.1. Antecedentes internacionales

3.1.1. Simplificación de la incorporación de pacientes

HCA Healthcare es un proveedor de atención líder con más de 180 hospitales y 2000 centros de atención en los Estados Unidos, que atiende a aproximadamente 35 millones de pacientes cada año.

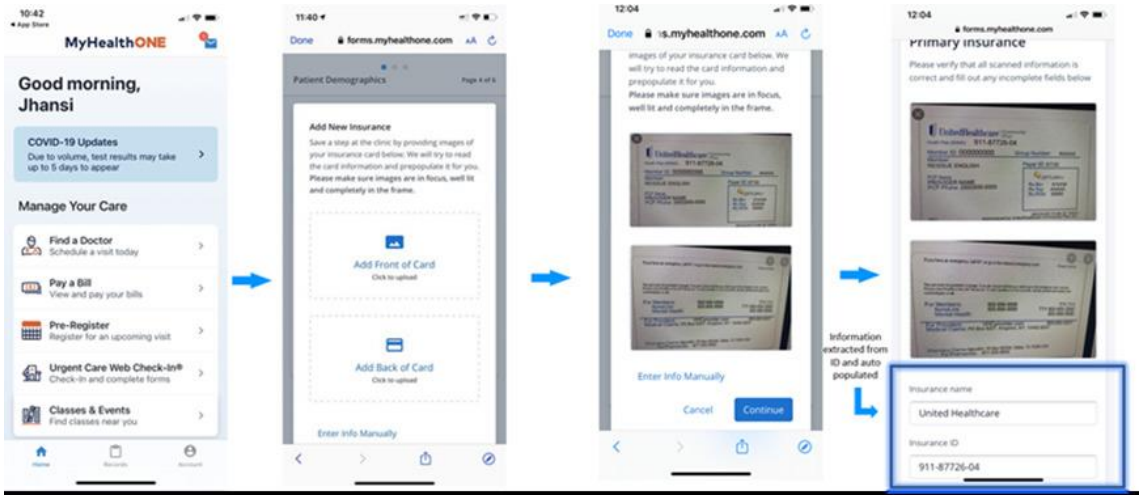
HCA Healthcare utiliza Form Recognizer para simplificar y mejorar la experiencia de visita del paciente y reducir el tiempo administrativo dedicado a ingresar datos duplicados en los sistemas del centro de atención.

Usando la aplicación MyHealthOne de HCA en un smartphone, los pacientes pueden tomar una foto de su identificación y tarjeta de seguro, que luego envían a HCA antes de acudir a una cita. Form Recognizer se encarga de identificar los campos que HCA quiere capturar y los registra automáticamente. Esto permite que HCA procese registros de pacientes sin ingreso manual de datos y, lo que es más importante, en el mundo actual, sin contacto físico. (Healthcare, 2021)

En la figura 21 se muestra el proceso de extracción de datos.

Figura 21

Funcionamiento de la aplicación MyHealthOne



Nota. Figura extraída de los casos de éxito de Microsoft Form Recognizer. Customer Microsotf, 2021 (https://customers.microsoft.com/es-es/story/1404891793134114534-hca-healthcare-healthcare-provider-azure).

3.1.2. Acelerando el tiempo de las conclusiones

Durante años, las operaciones canadienses de Chevron han extraído manualmente datos clave de los informes de perforación y finalización, un proceso engorroso, lento y propenso a errores. La compañía ahora usa Form Recognizer, parte del Servicio de Inteligencia Artificial Aplicada de Azure de Microsoft, junto con la plataforma de Automatización de Procesos Robóticos UiPath para extraer datos automáticamente y moverlos a sistemas de back-end para su análisis. Los expertos en la materia ahora tienen más tiempo para concentrarse en actividades de mayor valor, la información fluye más rápido y se acelera el control operativo, lo que permite a la empresa analizar su negocio con mayor rapidez, precisión y profundidad. (Chevron, UiPath, 2019)

3.1.3. Ahorrando tiempo en cada transacción

Standard Bank es el banco más grande de África por activos. Con sede en Johannesburgo, Sudáfrica y con más de 150 años de historia, Standard Bank está profundamente involucrado en el comercio tanto en el continente africano como más allá.

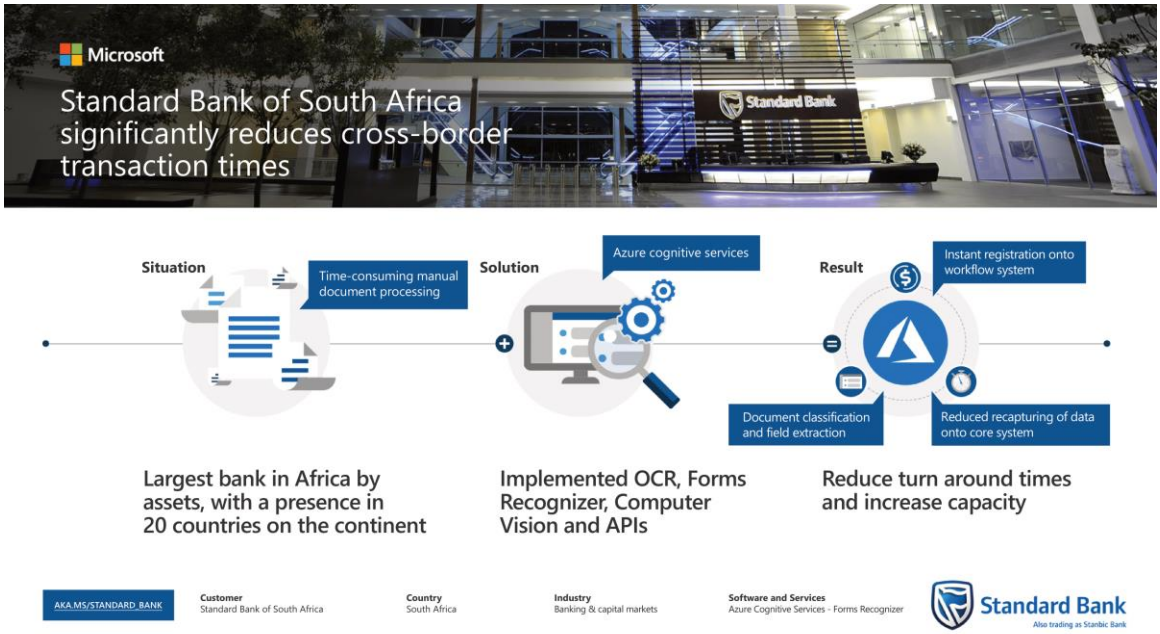
Pero cuando la debida diligencia manual en las transacciones transfronterizas comenzó a consumir demasiado tiempo del personal, el banco decidió que necesitaba un nuevo camino a seguir. Con Form Recognizer de Azure Cognitive Services, Standard Bank ahora puede reducir significativamente el tiempo de procesamiento y registro de pagos transfronterizos.

Standard Bank de South Africa utiliza Form Recognizer para automatizar transacciones internacionales complejas, mejorando el servicio al cliente y reduciendo significativamente los tiempos de procesamiento y los costos operativos. (Standard Bank of South Africa, 2021)

En la figura 22 se muestra el flujo de la solución implementada por Standard Bank.

Figura 22

Flujo de solución implementado por Standard Bank



Nota. Figura extraída de los casos de éxito de Microsoft Form Recognizer. Customer Microsotf, 2021 (https://customers.microsoft.com/es-es/story/1395059149522299983-standard-bank-of-south-africa-banking-capital-markets-azure-en-south-africa).

3.2. Antecedentes nacionales

3.2.1. Transformando recursos de manera sostenible

Minsur opera la mina de estaño más grande del hemisferio occidental en Puno, Perú. Para hacer que el proceso de inspección visual en el proceso de tratamiento de agua en la mina San Rafael sea más eficiente y seguro, fue creada una solución de IA que combina Microsoft Azure Video Analytics y Cognitive Services.

Este cambio significa ahorros financieros, ya que los funcionarios no tienen que monitorear las 24 horas del día para un control más preciso y directo. (Minsur S.A., 2019)

3.2.2. Automatización de procesos de facturación y contratos

EY (Ernst & Young) es una compañía líder global en servicio de Auditoría, Consultoría, Transacciones & Finanzas Corporativas con presencia en Perú.

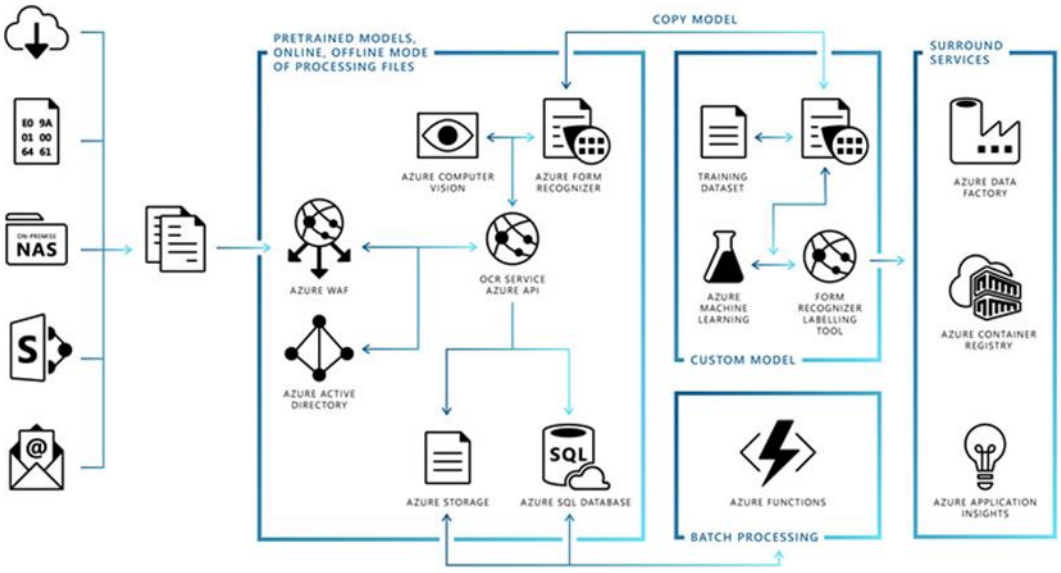
EY ayuda a los clientes a extraer datos de contratos y facturas utilizando Form Recognizer, ahorrando cientos de miles de horas de trabajo manual al mismo tiempo.

Parassery, gerente de producto, dice: "Si cada contrato tarda un promedio de tres minutos en analizarse, esto tiene el potencial de ahorrar 1,5 millones de minutos, o 250.000 horas, de trabajo manual". (EY, 2021)

En la figura 23 se muestra la arquitectura de Azure usada por EY.

Figura 23

Arquitectura Azure usada por EY



Nota. Figura extraída de los casos de éxito de Microsoft Form Recognizer. Customer Microsotf, 2021 (https://customers.microsoft.com/es-es/story/1404985164224935715-ey-professional-services-azure-form-recognizer).

3.3. Objetivos

3.3.1. Objetivo general

Agilizar las declaraciones aduaneras de mercancías a SUNAT automatizando el llenado de datos de diferentes documentos aduaneros en el área de servicio al cliente y producción, logrando mayor efectividad en la operación y optimizando tiempos de atención que permitirá atender a nuevos clientes.

3.3.2. Objetivos específicos

- Minimizar errores en el proceso de digitalización.
- Optimizar tiempos de digitalización de documentos como Facturas y Documentos de embarque.
- Clasificar documentos automáticamente, permitiendo el usuario realizar una carga masiva de documentos sin tener que clasificarlo individualmente.

CAPÍTULO IV

4. Discusión y Aportes

Los servicios Computer Vision y Form Recognizer de Microsoft Azure no son los únicos en el mercado que brindan las funcionalidades explicadas y usadas en el presente informe. También existen soluciones como Google Document AI y Amazon Textract con los cuales se pueden llegar resultados similares. En este caso se eligió Form Recognizer de Azure por la facilidad de uso de su plataforma, velocidad, costos y mejoras constantes que vienen realizando.

En la tabla 9 se muestra un comparativo de los 3 servicios mencionados después de realizar algunas pruebas analizando algunas facturas comerciales.

Tabla 9

Comparativo de servicios similares con Machine Learning

Servicio	Cuadro delimitador con OCR	5 muestras		10 muestras	
		Tiempo de entrenamiento tomado	Precisión	Tiempo de entrenamiento tomado	Precisión
Amazon	No	52 minutos	3 regiones no identificadas	1 hora 3 minutos	3 regiones no identificadas
Google	No	-	-	1 hora 14 minutos	Completo
Microsoft	Si	3 segundos	1.5 regiones no identificadas	5 segundos	1.5 regiones no identificadas

Azure Form Recognizer en solo tres segundos puede entrenar un modelo con un grado de precisión muy alto. Cuando se trata de precisión, Google es 100% preciso. Sin embargo, la capacitación lleva un poco más de una hora y requiere al menos diez documentos de muestra, en comparación con Microsoft o Amazon, que solo necesitan un mínimo de cinco muestras. En general, Microsoft Azure es el mejor de las opciones, por tal motivo, fue la que se usó para llevar a cabo este proyecto.

5. Conclusiones y Recomendaciones

- La integración de técnicas de Inteligencia Artificial como Machine Learning al módulo de "Gestión Documentaria" dieron como resultado una clasificación automática de documentos y la extracción de datos específicos evitando la digitalización de estos por parte del personal humano, reduciendo tiempos de atención de hasta 75% y errores de digitalización que conllevaban a rectificaciones y multas.
- La implementación desarrollada da pase a la automatización de próximos procesos como regularización de facturas del área administrativa que actualmente lo realiza el personal humano y se viene analizando para su próxima automatización.
- El servicio por usar (Microsoft Form Recognizer, Amazon Textract, Google Document AI) depende de las necesidades del negocio, ya que los formatos y datos de documentos aduaneros son distintos a otros tipos de documentos. Se recomienda primero analizar la problemática que se quiere resolver y que servicio brinda la solución que más se adapte o cubra las necesidades.

BIBLIOGRAFÍA

Chevron, UiPath. (6 de Mayo de 2019). *Microsoft Customers*. Obtenido de https://customers.microsoft.com/es-es/story/chevron-mining-oil-gas-azure-cognitive-services

CLI, G. A. (s.f.). *CLI*. Obtenido de https://www.cli.com.pe/nosotros.php

EY. (26 de Agosto de 2021). *Microsoft Customers*. Obtenido de https://customers.microsoft.com/es-es/story/1404985164224935715-ey-professional-services-azure-form-recognizer

Healthcare, H. (20 de Agosto de 2021). *Microsoft Customers*. Obtenido de https://customers.microsoft.com/es-es/story/1404891793134114534-hca-healthcare-healthcare-provider-azure

Microsoft. (s.f.). *Computer Vision*. Obtenido de Microsoft Azure: https://azure.microsoft.com/es-es/products/cognitive-services/computer-vision

Microsoft. (s.f.). *Form Recognizer*. Obtenido de Microsoft Azure: https://azure.microsoft.com/es-es/products/form-recognizer/

Minsur S.A., A. (7 de Febrero de 2019). *Microsoft Customers*. Obtenido de https://customers.microsoft.com/es-es/story/minsur-artifical-intelligence-azure-cognitive-services-machine-learning-mining-oil-arkano-peru-es

Standard Bank of South Africa, M. (28 de Julio de 2021). *Microsoft Customers*. Obtenido de https://customers.microsoft.com/es-es/story/1395059149522299983-standard-bank-of-south-africa-banking-capital-markets-azure-en-south-africa

FORMATO DE AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL UNAM

Formulario de autorización para la publicación electrónica en la página web del Repositorio Institucional de la Universidad Nacional de Moquegua, de conformidad con la Ley N° 30035 Ley que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto y Resolución de Comisión Organizadora N° 041-2022-UNAM que aprueba el Reglamento que regula el Repositorio Institucional Digital de la UNAM.

1. DATOS DEL AUTOR

Nombres y Apellidos : Darwin Daniel Enrique Apaza Flores
DNI/Carnet de Extranjería/Pasaporte N° : 75005944
Correo electrónico : darwindeaf@gmail.com
Teléfono/Celular : 931071062

2. GRADO, TITULO O ESPECIALIZACIÓN (Marque con una X)

Bachiller	Título Profesional	Magister	Doctor	Segunda Especialidad
	X			

3. TIPO DE PUBLICACIÓN (Marque con una X)

a) Articulo ( )
b) Tesis de pregrado ( )
c) Tesis de Maestría ( )
d) Tesis de Doctorado ( )
e) Monografía (**X**)
f) Reporte de caso científico o hallazgos ( )
g) Libro ( )
h) Parte de un libro ( )
i) Revisión ( )
j) Conferencia ( )
k) Presentación ( )
l) Artículo preliminar ( )
m) Articulo antes de ser publicado ( )
n) Artículo de prensa ( )
o) Documentos internos ( )
p) Imagen ( )
q) Documentos técnicos ( )
r) Ponencias presentadas en eventos científicos (Jornada, Congreso, Simposio Conferencias, etc.) ( )
s) Informes de suficiencia profesional ( )
t) Patentes producidas por académicos/as ( )

04

4. TIPO DE TRABAJO DE INVESTIGACIÓN (Marque con una X)

a) Tesis ( )
b) Trabajo de Investigación ( )
c) Trabajo de suficiencia profesional ( )
d) Trabajo académico (**X**)

5. DATOS DE LA PUBLICACION

a) Título : IMPLEMENTACIÓN DE LECTURA DE DOCUMENTOS DIGITALES CON MACHINE LEARNING EN LA EMPRESA CLI GESTIONES ADUANERAS S.A.
b) Campo del conocimiento OCDE (Indicar el código de clasificación OCDE de la obra : https://catalogos.concytec.gob.pe/vocabulario/ocde_ford.html) : 2.02.04
c) Fecha de publicación : 2022

6. DATOS DEL ASESOR:

a) Nombres y Apellidos : Victor Yana Mamani
b) DNI/Pasaporte/Cédula : 02437887
c) ORCID : https://orcid.org/0000-0003-0982-2353

FIRMA : .....

DOCUMENTO DE IDENTIDAD : 75005944
NOMBRES Y APELLIDOS : Darwin Daniel Enrique Apaza Flores

LICENCIAS

Licencia Estándar

Bajo los siguientes términos, autorizo el depósito de mi trabajo de investigación en el Repositorio Institucional Digital de la UNAM.

Con la autorización de depósito da mi trabajo de investigación, otorgo a la Universidad Nacional de Moquegua, una Licencia No Exclusiva para reproducir. distribuir, comunicar al público, transformar (únicamente mediante su traducción a otros idiomas) y poner a disposición del público mi trabajo de investigación (incluido el resumen), en formato físico o digital, en cualquier medio, conocido o por conocerse, a través de los diversos servicios provistos por la Universidad, creados o por crearse, tales como el Repositorio Institucional Digital de la UNAM, Colección de Tesis, entre otros, en el Perú y en el extranjero, por el tiempo y veces que considere necesarias, y libre de remuneraciones.

En virtud de dicha Licencia, la UNIVERSIDAD NACIONAL DE MOQUEGUA, podrá reproducir mi trabajo de investigación en cualquier tipo de soporte y en más de un ejemplar, sin modificar su contenido, solo con los propósitos de seguridad, respaldo y preservación.

Declaro que el trabajo de investigación es una creación de mi autoría y exclusiva titularidad, o coautoría con titularidad compartida, y me encuentro facultado a conceder la presente licencia y, asimismo, garantizo que dicho trabajo de investigación no infringe derechos de autor de terceras personas.

La UNIVERSIDAD NACIONAL DE MOQUEGUA consignara el nombre del/los autor/es del trabajo de investigación, y no le hará ninguna modificación más de la permitida en la presente licencia.

Autorizo el depósito (marque con una X).

(**X**) **SI autorizo que se deposite inmediatamente.**

FIRMA : .....



DOCUMENTO DE IDENTIDAD : 75005944
NOMBRES Y APELLIDOS : Darwin Daniel Enrique Apaza Flores
FECHA : 05 de Octubre de 2022