

Reto Rappi: MA2001S

Alumno:	Contacto:	Escuela:
Mariana Rizo	A01423985@itesm.mx	ITESM
María Valentina García	mv.garcia@uniandes.edu.co	UniAndes
Ana María Montaña	am.montana10@uniandes.edu.co	UniAndes
Alberto Cortés	A01635875@itesm.mx	ITESM
Alejandro Palomino	a.palomino@uniandes.edu.co	UniAndes
Juan Sebastian Rodríguez	js.rodriguezp@uniandes.edu.co	UniAndes

1. Definición del proyecto

1.1 Definición del problema

A mediados de enero del 2020, Rappi comenzó a promocionar una tarjeta de crédito en países como Colombia, México, Brasil y Perú. Lo diferente e innovador de este producto radica en que se puede obtener mediante un proceso completamente digital. Dentro de este, existe el proceso Know Your Customer (KYC), el cual consiste en la verificación de identidad de los posibles clientes. La apuesta de Rappi es poder tener procesos rápidos, que el mismo día que la persona manda su solicitud se le pueda tener respuesta. Puesto que, la empresa no cuenta con cajeros que puedan hacer el análisis de la información del cliente, sino que se pueda automatizar con AI. Y así, el cliente en unos 10 minutos ya tendrá su tarjeta digital.

Actualmente, la empresa Rappi cuenta con un proveedor que hace tareas de Object Character Recognition (OCR), el cual consiste en traducir la información de documentos a datos digitales que se puedan usar. Sin embargo, la empresa desea crear un mecanismo interno que les permita mayor flexibilidad y reducción de costos. El reto de OCR es uno de los pasos que la empresa toma dentro del proceso digital para analizar, aprobar y entregar las tarjetas Rappi a sus clientes. Para esto, la empresa desea extraer información como nombre, apellidos, domicilios, fecha de nacimiento y número de identificación (CURP). Este reto, a su vez, contiene 3 niveles:

1. Nivel Básico de la identificación
2. Generalización de identificaciones
3. Generalización a otros documentos (comprobante de domicilio, etc.)

Sin embargo, a un plazo más largo, se espera que este proyecto se pueda expandir a distintos países y que el mecanismo interno sea capaz de llevar a cabo las tareas de OCR sin importar de qué país viene la identificación.

1.2 Descripción de la solución y los algoritmos

El reconocimiento de imágenes, en este caso, resulta ser la respuesta a la problemática presentada por la empresa Rappi. Usando visión computacional e inteligencia artificial, se pueden automatizar, y por ende economizar, algunos procesos que la empresa desea llevar a cabo. Para hacer esto, será necesario contar con una base de datos lo suficientemente grande como para poder entrenar al modelo. En este caso decidimos optar por usar un método de aumentación de datos, con el cual podremos ingresar las identificaciones oficiales de los países respectivos y generar imágenes con un grado de brillo, saturación, ruido, etc., diferentes. Todas estas imágenes, tanto originales como generadas, pasarán por un proceso de etiquetación, en el cual se etiquetará la ubicación del nombre, apellidos, dirección, sexo, número de identificación (CURP) y fecha de nacimiento. Un ejemplo de esto esto se ve en la Figura 1.



Figura 1. Izquierda: Ejemplo de etiquetado de datos para un INE Mexicano. En este caso se pueden identificar todas las etiquetas mencionadas en el texto. Centro: Ejemplo de etiquetado en una cédula colombiana para la cara frontal de la cédula de ciudadanía, de la cual se identifican el número de identificación, nombre y apellidos. Derecha: Ejemplo de etiquetado de datos para una cédula colombiana por la cara trasera del documento. En este caso, solo se etiquetan la dirección, el sexo y la fecha de nacimiento.

Seguido de la parte de preparación de datos, se entrenarán estos datos usando la técnica de transferencia de datos. Esta técnica aprovecha una de las características de las redes neuronales. Ya que las redes neuronales están construidas en capas, es posible insertar conocimiento entre cualquiera de las capas, de tal manera que la red neuronal se pueda entrenar de manera más efectiva, ya que una vez que se ingresan datos nuevos tendrá más información para el entrenamiento. Esto presenta varias ventajas como mayor precisión, capacidad y una tasa de convergencia mejor (Ruder, 2017).

Para la fase de predicción se usará una implementación que únicamente usa una red neuronal en toda la imagen, YOLO: You Only Look Once, la cual cuenta con varias ventajas como la velocidad del modelo, la alta precisión y la alta capacidad de aprendizaje, el cual se ve afectado por el contexto global (Sachan, 2019). Para ello, se usarán los archivos de texto plano (.txt) generados en la primera etapa del proyecto, los cuales contienen las etiquetas y las posiciones de ellas en las imágenes. La estructura de este modelo está basada en CNN (Convolutional Neural Networks), la cual es capaz de predecir la probabilidad de que un objeto, en este caso caracteres, pertenezcan a cierta clase y dónde en la imagen se encuentran estas clases.

En la Figura 2, se puede observar cómo funciona, paso a paso, YOLO. El primer paso es la división de la imagen en pequeñas partes, de tal manera que se forme una matriz SxS. Después de esto se llevan a cabo dos procesos. Uno de estos procesos consiste en generar un mapa de probabilidad para las clases con las que se han entrenado los datos. El otro proceso define cuántos objetos hay en la imagen delimitando cuadrados, lo cual es un proceso similar al que se llevó a cabo en el etiquetado manual. Con la información de estos procesos, YOLO puede predecir la probabilidad de las clases y los objetos que pertenecen a ella (Karimi, 2021).

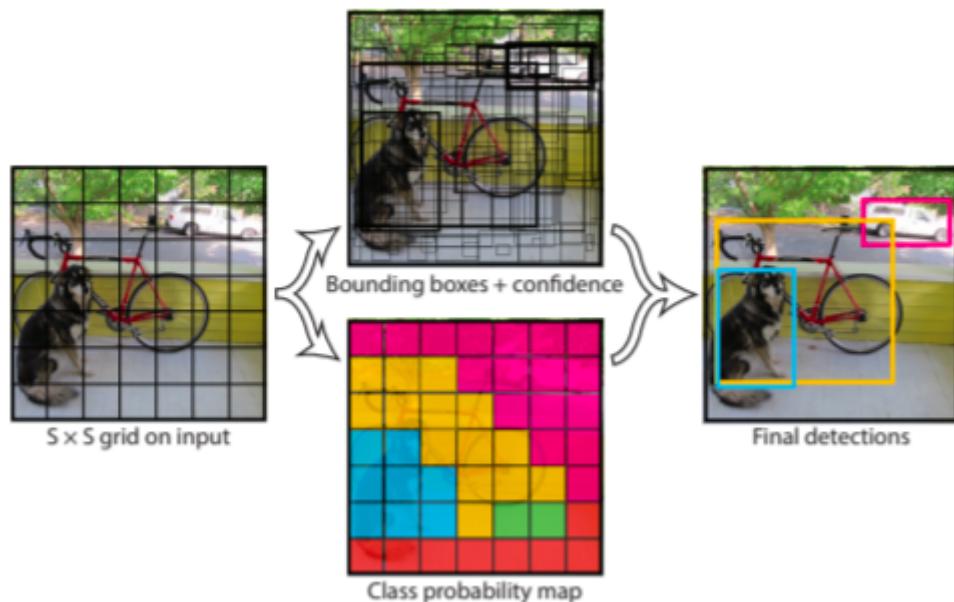


Figura 2. Funcionamiento de YOLO, explicado con una imagen en la que se clasifican tres elementos.

1.3 Pré-requisitos

Los datos que deberán ser proporcionados por el cliente se pueden encontrar en la identificación oficial. En caso de ser una identificación oficial mexicana sólo se necesitará una captura de la parte frontal, y, para la identificación oficial colombiana, se necesitarán capturas de ambos lados. La información que se extraerá de estas será la siguiente:

- Nombre(s)
- Apellido(s)
- Domicilio
- Fecha de nacimiento
- Sexo
- Número de identificación

Estas clases estarán indexadas numéricamente mediante un ID único para cada una de ellas según lo muestra la siguiente tabla.

ID	Clase
0	Nombre(s)
1	Apellido(s)
2	Domicilio
3	Fecha de nacimiento
4	Sexo
5	Número de identificación

Tabla 1. Indexación numérica de las seis clases creadas

Para el caso colombiano, y ya que las cédulas de ciudadanía no cuentan con la dirección, se toma el lugar de expedición de la cédula dado que, aunque no se garantiza que la persona viva en el mismo lugar en el que realizó el trámite, es lo más cercano al dato de dirección en el documento de identificación colombiano.

En cuanto al formato exacto de las imágenes que debe proporcionar el cliente, deben ser imágenes en formato .png o .jpg. Las imágenes deben estar en formato horizontal y los datos en la imagen deben ser legibles. Finalmente, el documento debe verse completo, es decir, no debe estar cortado. En la Figura 2. se muestran algunos ejemplos de imágenes no permitidas.



Figura 3. Ejemplos de imágenes no permitidas como entrada. Izquierda: Imagen rotada, es decir, no está en formato horizontal. Centro: Imagen recortada en la cual no se ven completos los datos. Derecha: Imagen con los datos no legibles.

2. Evaluación de los experimentos

Se utilizó un modelo basado en YOLOv5 caracterizado por ser simple, liviano en complejidad, interactivo y sumamente eficaz. El principal problema con este modelo es que no es completamente independiente, sino que su código está basado en la implementación de detección de imágenes de Roboflow.

2.1 Descripción de los datos de entrada

Para este proyecto se utilizaron 49 identificaciones oficiales de Colombia y 40 de México. Estas imágenes de las identificaciones pasaron por un preprocesamiento en el cual se aumentaron, usando la librería de *AugLy*, con lo cual se logró poblar un dataset de 333 imágenes, las cuales se dividieron en 291 de entrenamiento, 28 de validación y 14 de testeo. Es importante destacar que se descartaron algunas imágenes aumentadas debido a que no cumplían con los requisitos de la sección 1.3.

Justo después de completar el dataset se realizó el etiquetado, el cual se llevó a cabo en dos etapas:

- Etapa 1: Para las imágenes originales, y hasta 139 imágenes aumentadas de todo tipo (tamaño, posición, color, etc), se realizó un etiquetado manual.
- Etapa 2: Se generaron imágenes aumentadas que no fueran transformadas en términos de posición o tamaño con el fin de agilizar el etiquetado. Esto último dado que, para este grupo, las coordenadas y dimensiones de las cajas de etiquetas son exactamente iguales a las del documento original, el cual fue clasificado manualmente en la etapa 1.

El etiquetado de las imágenes se llevó a cabo usando los archivos de Ybat - YOLO BBox Annotation Tool (drainingsun, 2019). Esta herramienta permite que se seleccionen los cuadros en dónde se encuentra cada clase a partir de un documento en texto plano que indique las clases y almacena las coordenadas, alto y ancho en píxeles de cada uno de estos bloques en archivos de texto plano. Estos archivos de texto plano que contienen la información de la ubicación de cada clase, anteriormente mencionadas en la sección 1.3, se usan en conjunto con las imágenes para entrenar el modelo de YOLO.

2.2 Metodología de la evaluación

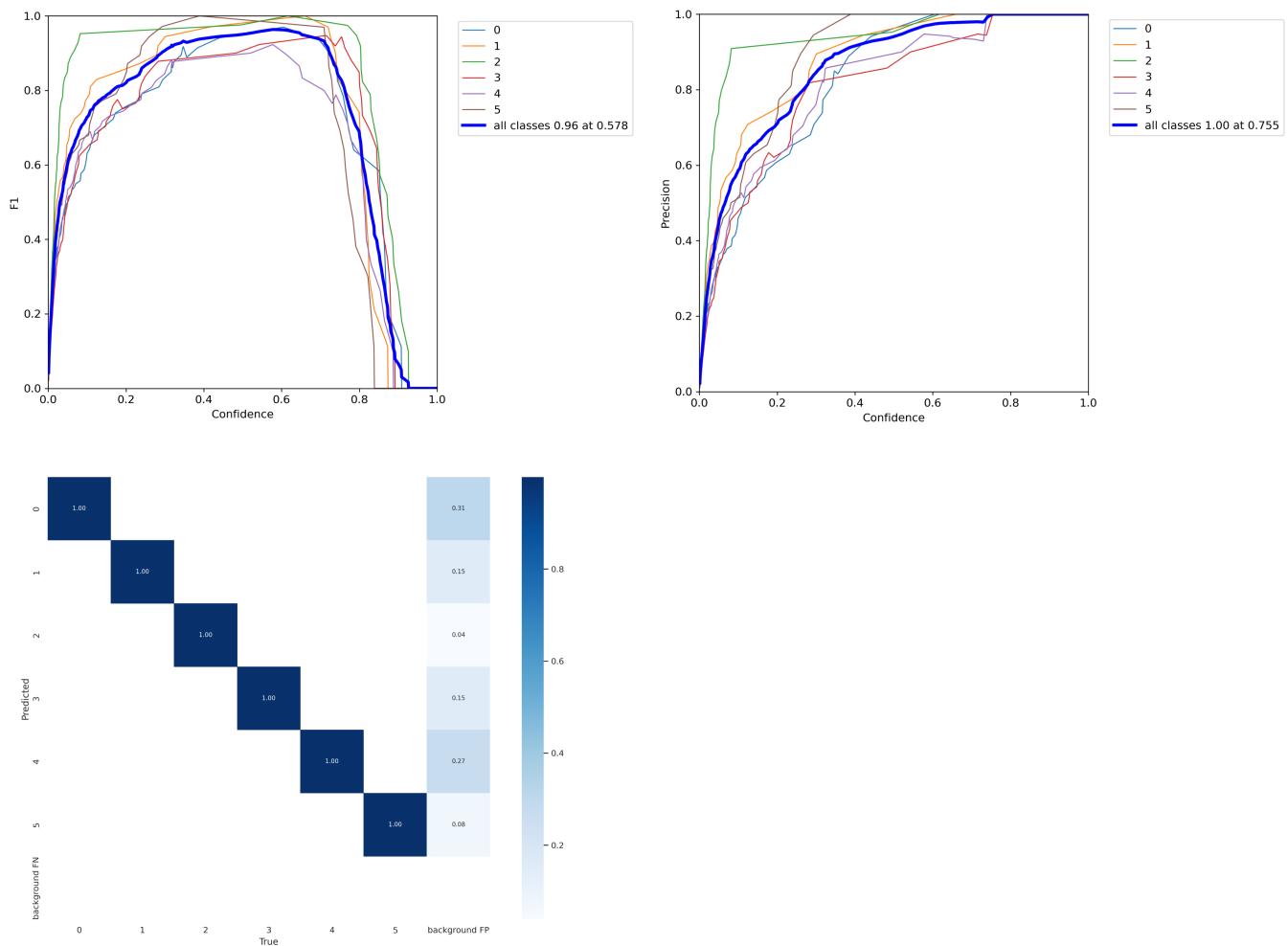
Las métricas que se utilizarán para evaluar el modelo son F1, precisión y la matriz de confusión. Se selecciona el primer estadístico ya que este contiene información tanto de la precisión como del recall, obteniendo así una observación sobre qué tan exhaustivo es el modelo para reconocer los datos de interés en diferentes credenciales así como qué tan bien lo hace. El segundo estadístico se usa porque este aporta información plana y simple sobre la probabilidad del modelo para identificar los datos en las credenciales. El último se utiliza para tener una perspectiva más general del desempeño del modelo a lo largo de todas las clases y cómo es que está caracterizando cada una de ellas.

2.3 Resultados

A continuación se puede ver un resumen de los resultados obtenidos por el modelo.

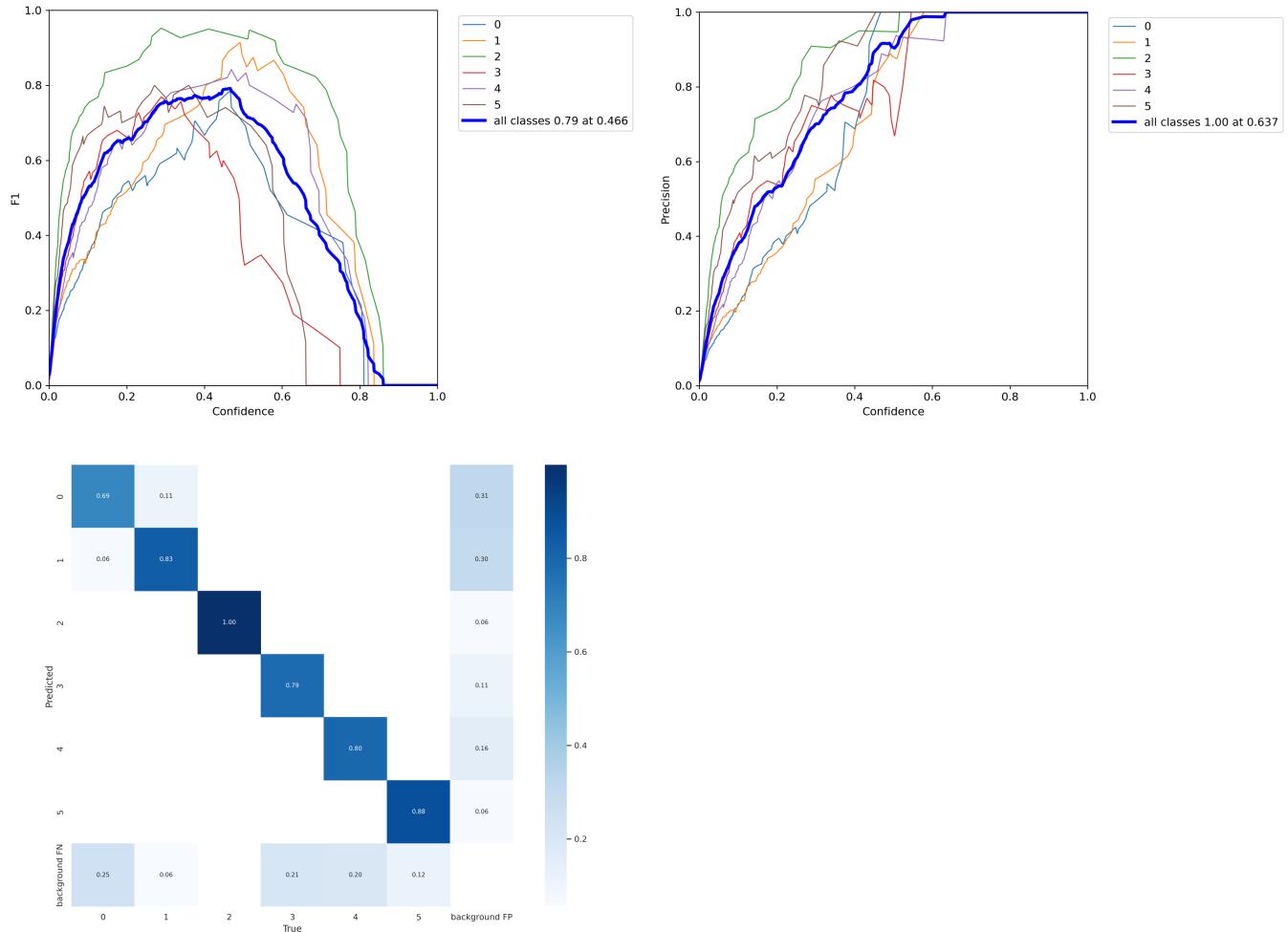
Se realizó un *grid search* con los meta-parámetros del número de épocas y el tamaño del *batch*, estos arrojaron que la configuración óptima del modelo es con X épocas y Y datos por batch. Luego de utilizar esta configuración se entrenó el modelo, presentando resultados para 4 batches. Los resultados fueron los siguientes:

- Epoch = 100, Batch = 16.
 - F1 de 0.96 con una confiabilidad de 0.578
 - Precisión de 1.00 con una confiabilidad de 0.755
 - Diagonal principal de la matriz de confusión con valores de 1.00 para cada clase.
 - Precisión de 0.9732 al final del batch.

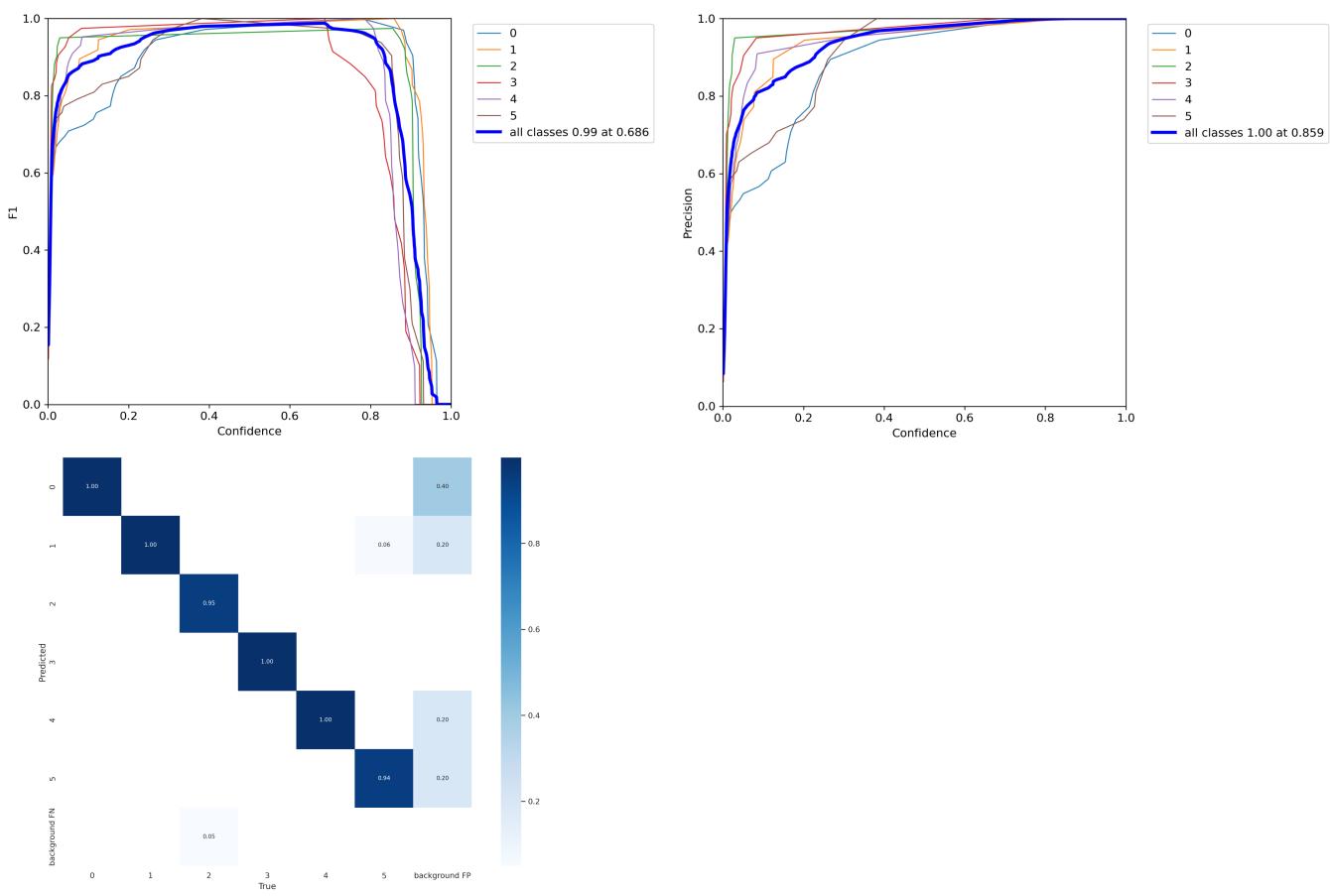


- Epoch = 100, Batch = 32.
 - F1 de 0.79 con una confiabilidad de 0.466
 - Precisión de 1.00 con una confiabilidad de 0.637

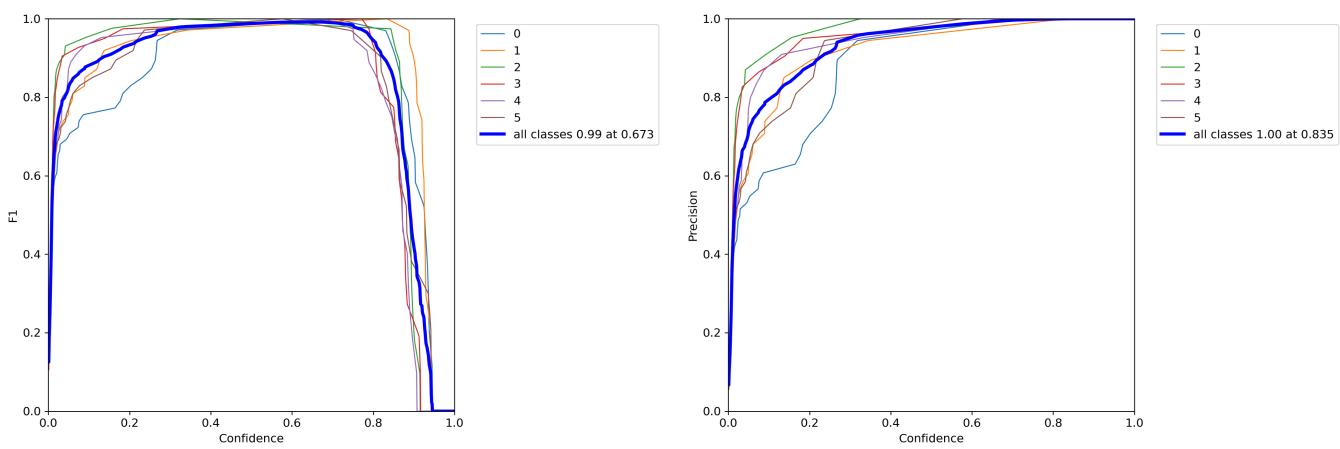
- Diagonal principal de la matriz de confusión con valores altos pero no ideales.
- Précision de 0.913 al final del batch.

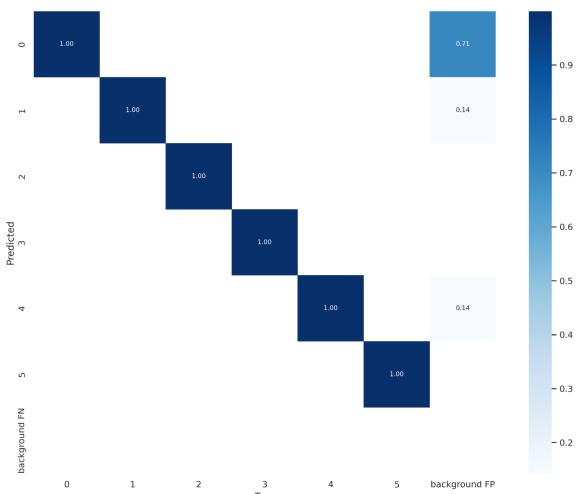


- Epoch = 200, Batch = 16.
 - F1 de 0.99 con una confiabilidad de 0.686
 - Precisión de 1.00 con una confiabilidad de 0.859
 - Diagonal principal de la matriz de confusión con valores mayormente de 1.00.
 - Précision de 0.992 al final del batch.



- Epoch = 200, Batch = 32.
 - F1 de 0.99 con una confiabilidad de 0.673
 - Precisión de 1.00 con una confiabilidad de 0.835
 - Diagonal principal de la matriz de confusión con valores de 1.00.
 - Précision de 0.995 al final del batch.





En estos últimos resultados se observa un incremento notable en el desempeño del modelo, alcanzando una precisión cercana a 1.00 y una matriz de confusión casi perfecta de igual manera. Una vez que el modelo alcanzó este nivel se llevaron a cabo predicciones con datos nunca antes vistos por el modelo, estas presentaron los siguientes resultados.



3. Trabajo futuro

Uno de los retos más grandes a los que se ve enfrentado este proyecto es la recolección de datos. Dado que los datos que se tendrían que manejar en esta implementación de OCR son privados y sensibles, será necesario contar con un sistema que asegure la seguridad de los datos a los clientes. Durante esta semana, trabajamos con un conjunto de datos relativamente pequeño, pero se logró potencializar al implementar diferentes métodos de aumentación. Consideramos que estos métodos de aumentación pueden ser útiles a futuro, pero su uso tiene límites de eficiencia en el resultado final, debido a posibles sesgos por falta de variabilidad en los datos de entrenamiento y testeo. Es por esto que el primer trabajo a futuro es lograr construir un dataset lo suficientemente robusto como para poder entrenar y testear nuestro modelo con una gran variedad de imágenes y a la vez garantizando la privacidad y seguridad de los datos recolectados.

Adicionalmente, se podría hacer la verificación y obtención de información adicional de las cédulas de las personas o de otros documentos legales a los que una organización pueda tener acceso. También, como trabajo futuro es posible incluir la generalización a otro tipo de documentos como lo son pasaportes, documentos de extranjería, licencias, documentos del domicilio o demás que puedan ser de interés para la organización. Para ello, se requeriría de un mayor tiempo de ejecución del proyecto con el fin de hacer un ejercicio en el que se pueda obtener mayor cantidad de información o que se pueda obtener datos en la ejecución de alguna actividad organizacional.

A su vez, se podrían implementar algoritmos para el pre-procesamiento de los datos, de tal forma que más formatos de las imágenes de entrada sean permitidos. Un ejemplo de esto sería poder recibir imágenes rotadas y en formato .pdf. Todo esto con la finalidad de facilitar y economizar procesos que utilicen imágenes como fuente de información.

Además de todo esto se podría realizar un análisis de fuerzas de Porter, esto para comprender mejor el posible impacto de la automatización, e involucramiento de Machine Learning en el industrial de obtención de datos, este análisis tendría la siguiente estructura:

- El primer aspecto de Porter es denominado Nuevos Entrantes. Este aspecto hace referencia a la posibilidad de entrada de nuevas compañías que sean competencia en el mercado. Se podría decir que la implementación del algoritmo permitiría fortalecer a la organización en este aspecto ya que se tendrían tareas automatizadas que facilitan soluciones en el ejercicio de sus funciones. Sin embargo, se debe tener en cuenta que el código implementado podría, en un futuro, tener la capacidad de procesar cantidades de datos diferentes con un porcentaje de asertividad más alto tal que la información proporcionada sea de mayor valor para la organización.
- El segundo aspecto de Porter es denominado proveedores. Este se refiere al poder que posee la organización para la negociación con los proveedores. En este caso, dado que se habla de la implementación del algoritmo, se debe tener especial

cuidado con quien provee la información para poder ponerlo en funcionamiento. Se le recomienda a la organización revisar este aspecto ya que el manejo de esta información trae consigo requisitos legales que se deben tener en cuenta con el fin de evitar conflictos futuros.

- El tercer aspecto de Porter son los clientes. Este se refiere al poder que posee la organización para negociar con los clientes. En este caso, la implementación trae consigo facilitar tareas en la estructura organizacional interna, por lo que la compañía se puede comprometer a otorgar soluciones más completas o con análisis adicionales en sus tareas ya que algunas de sus funciones fundamentales se encontrarán automatizadas.
- Por último, se tiene el aspecto denominado sustitutos. Este aspecto de Porter se basa en la amenaza de productos sustitutos. En este aspecto, la organización debe prestar especial atención teniendo en cuenta que, actualmente, es muy común el desarrollo de algoritmos de Machine Learning que realicen las tareas de automatización. Por lo tanto, se busca, como se dijo anteriormente, hacer una continua actualización y mejora del código con el fin de evitar la entrada de productos sustitutos que generan costos elevados adicionales para la compañía.

4. Conclusión

El desarrollo de modelos de inteligencia artificial dejó de ser un ejercicio estrictamente académico cuando este alcanzó un potencial realmente aprovechable en la industria, siendo tan natural el desarrollo de estos modelos en empresas nacidas en el internet cuyo negocio es completamente tecnológico, una de ellas siendo Rappi.

Si bien las necesidades tecnológicas de cada empresa son diferentes, los acercamientos de inteligencia artificial también son amplios y es a través de estos que se pueden llevar a cabo tareas que si bien son simples de realizar por un operador humano también le consumen tiempo que podría ser aprovechado en tareas de mayor valor para la empresa.

Siguiendo esta filosofía es que se logró:

- Entrenar un modelo basado en una arquitectura ya existente con los datos específicos de este problema.
- La identificación de datos dentro de credenciales de identificación Mexicanas y Colombianas con un alto nivel de precisión para todas las clases, encontrando así que sin importar el formato de la credencial es posible extraer los datos necesarios para resolver las necesidades de Rappi.

Si bien experimentos iniciales utilizando otra arquitectura de red no demostraron mayor efectividad, puesto que contaban con valores de pérdida cercanos a 1 y precisión cercanas a 0, luego de varias épocas de entrenamiento, realizar un cambio de arquitectura hacia la que se terminó reportando demostró el valor que había en el conjunto de datos aumentando y etiquetado.

Una vez que se puede aprovechar el modelo para etiquetar datos nuevos como se observa en la presentación de resultados es que se demuestra la viabilidad de la solución.

Bibliografía

- drainingsun. (2019, 12 1). *Ybat - YOLO BBox Annotation Tool*. GitHub.
<https://github.com/drainingsun/ybat>
- Karimi, G. (2021, 4 15). *Introduction to YOLO Algorithm for Object Detection*.
Section.
<https://www.section.io/engineering-education/introduction-to-yolo-algorithm-for-object-detection/>
- Ruder, S. (2017, 3 21). *Transfer Learning - Machine Learning's Next Frontier*.
ruder.io. <https://ruder.io/transfer-learning/>
- Sachan, A. (2019). *Zero to Hero: Guide to Object Detection using Deep Learning: Faster R-CNN, YOLO, SSD*. CV-Tricks.com.
<https://cv-tricks.com/object-detection/faster-r-cnn-yolo-ssd/>