

## **VELOX: Inteligencia artificial aplicada a las vistas. Una nueva PoC para la gestión interna de la Fiscalía de Estado de PBA**

### **VELOX: Artificial intelligence applied to views. A new PoC for the internal management of the PBA State Prosecutor's Office**

Mariano Cervellini<sup>1</sup>, Miguel Carbone<sup>1</sup>, Blas Szelagowski<sup>1</sup> Federico Gonzalez Oria<sup>1</sup>

<sup>1</sup> Fiscalía de Estado de la Provincia de Buenos Aires, Av. 1 esq 60, 1342

La Plata (1900), Argentina

{cervellini, carbone, bszelagowski, fpgonzalezoria}@fepba.gov.ar

#### **Resumen**

En el ámbito de la Fiscalía de Estado de la Provincia de Buenos Aires se desarrolló una prueba de concepto (PoC) que asiste a los abogados y abogadas en la confección de la vista fiscal que en sede administrativa emite el Fiscal de Estado, con carácter previo a decidir los expedientes en los que se reclama el pago de intereses por mora en la cancelación de facturas.

Se trata de una nueva aplicación de la tecnología desarrollada bajo la denominación VELOX, para pruebas realizadas previamente a fin de explorar técnicas de IA que permitan mejorar el desempeño del Organismo y lograr una eficaz defensa del patrimonio fiscal.

En este caso el abogado o abogada sube al sistema el expediente electrónico activando el proceso de búsqueda de datos guiada por la automatización del proceso, mediante la interacción con el usuario a través de la interfaz instrumentada en base a las preguntas y respuestas, con la intervención del usuario en caso de ser necesario.

Una vez identificados los datos disponibles en los documentos que integran el expediente, el sistema está en condiciones de incorporarlos a los modelos y/o de calcular mediante operaciones aritméticas nuevos datos necesarios.

A partir del cruce de la información recabada se constata si la factura ha sido pagada en término, si el reclamo es extemporáneo o si corresponde el pago de intereses por mora, luego de lo cual se genera la vista automáticamente.

No obstante el buen rendimiento de la solución aplicada al caso mediante la aplicación de técnicas de programación tradicionales, frente a variaciones detectadas en la estructura y contenido de los documentos de los que se extraen los datos, se decidió aplicar técnicas de machine learning aplicando aprendizaje supervisado, con resultados altamente promisorios.

**Palabras clave:** Automatización, Machine Learning, Clasificación, Aprendizaje supervisado.

#### **Abstract**

Within the scope of the State Prosecutor's Office of the Province of Buenos Aires, a proof of concept (PoC) was developed that assists lawyers in the preparation of the fiscal view issued by the State Prosecutor in administrative headquarters, prior to deciding the files in which the payment of interest is

claimed for late payment in the cancellation of invoices. It is a new application of the technology developed under the name VELOX, for tests previously carried out in order to explore AI techniques that improve the performance of the Agency and achieve an effective defense of fiscal assets. In this case, the lawyer uploads the electronic file to the system, activating the data search process guided by the automation of the process, through interaction with the user through the instrumented interface based on the questions and answers, with the intervention of the user if necessary.

Once the data available in the documents in the file have been identified, the system is able to incorporate them into the models and/or to calculate new necessary data by means of arithmetic operations. From the crossing of the information collected, it is verified if the invoice has been paid on time, if the claim is untimely or if the payment of interest for late payment corresponds, after which the *view* is automatically generated. Despite the good performance of the solution applied to the case through the application of traditional programming techniques, compared to variations detected in the structure and content of the documents from which the data are extracted, it was decided to apply machine learning techniques applying supervised learning, with highly promising results.

**Keywords:** Automation, Machine Learning, Classification, Supervised Learning.

## 1. Introducción

VELOX [1] es un prototipo que surge de una prueba de concepto (PoC) desarrollada durante el período 2020-2021<sup>1</sup> en el ámbito de la Fiscalía de Estado de la Provincia de Buenos Aires, como resultado de explorar técnicas de IA que permitieran mejorar el desempeño del Organismo para lograr una eficaz defensa del patrimonio fiscal [2].

---

<sup>1</sup> A partir del mes de julio de 2021 se instrumentó una prueba piloto por 6 meses. La misma fue evaluada de manera satisfactoria por la Subsecretaría de Relaciones Institucionales y Planificación Estratégica del Organismo en base a la teoría del cambio y orientado por las categorías de la cadena de resultados, con especial énfasis en la situación inicial, el producto y los resultados, como también en la percepción de actores involucrados respecto a algunos impactos producidos. En el “Reporte de Evaluación del Proyecto Velox” (Abril 2022) se menciona que “...La totalidad de las personas que expresaron su opinión en la encuesta de relevamiento sobre el uso de la herramienta fueron mujeres, con un promedio de edad de 39 años. VELOX fue calificado en su agilidad para el trabajo cotidiano con 8,6 puntos sobre 10. Las opiniones respecto a su uso y funcionamiento han sido muy positivas. La totalidad expresó que la herramienta funcionó bien para su cometido (“Excelente herramienta de trabajo” o “Resulta muy útil”). Del mismo modo, todas las personas entrevistadas y encuestadas expresaron que no presentó dificultades en su uso. También expresaron que el proceso de aprendizaje para incorporar esta nueva tecnología a las labores cotidianas se realizó sin dificultades;...” Y entre las conclusiones, se señala que “La utilidad de esta herramienta ha sido probada con altos niveles de satisfacción y utilidad en cuanto a su funcionamiento y uso, no encontrando quejas o reclamos respecto a esto último. También se destacó la facilidad para su aprendizaje y ejecución. Finalizada la etapa de prueba piloto, VELOX ya es considerado como necesario para hacer frente al cúmulo de trabajo...”



**Imagen 1.** Proyecto VELOX.  
Disponible en: <https://velox.ialab.fiscalia.gba.gov.ar/>

En base a la evidencia que muestra que aumentar la dotación de abogados no se traduce necesariamente en una mejora de la efectividad de los resultados en los tribunales [3], nuestro trabajo se focaliza en sumar tecnología para hacer frente al crecimiento sostenido de trabajo. De esa manera pretendemos liberar tiempo y recursos para el tratamiento de cuestiones de mayor complejidad por parte del plantel profesional experimentado del Organismo y aprovechar los frutos de años de trabajo y formación al servicio de la defensa pública.

Esas premisas, los promisorios resultados obtenidos y la participación activa de los propios agentes en un ámbito institucionalizado<sup>2</sup>, nos han permitido incorporar mejoras a VELOX, trasladar el concepto y proyectarlo hacia el resto de las dependencias.

Como consecuencia hoy contamos con una nueva PoC en marcha que asiste a los abogados y abogadas en la confección de la vista que en sede administrativa emite el Fiscal de Estado, con carácter previo a decidir los expedientes en los que se reclaman intereses por mora en el pago de facturas.

Siempre con el norte de continuar explorando nuevas técnicas que permitan superar los logros alcanzados<sup>3</sup>.

<sup>2</sup> Por Resolución RESO-2021-58-GDEBA-FED del 8/4/2021 el Fiscal de Estado se creó el “Laboratorio de Inteligencia Artificial para la Fiscalía de Estado” (FEPBA IALab), para el análisis y la incubación de soluciones de Inteligencia Artificial (IA) a partir de un modelo de innovación abierta que permita acelerar los procesos de investigación y desarrollo. Sin embargo, desde finales 2018 empezamos a vislumbrar el potencial de la inteligencia artificial y su impacto en el derecho, y su aplicación en las tareas del Organismo, lo que fue esbozado en la propuesta programática que se denominó “Fiscalía de Estado 2030”. Aún sin ese marco, nos propusimos continuar con las ideas allí postuladas, y a partir del año siguiente nos fijamos como metodología de trabajo realizar pruebas de concepto aplicando inteligencia artificial. La presente se trata de la quinta prueba de concepto realizada y será la segunda prueba piloto que el Laboratorio despliega.

<sup>3</sup> El Laboratorio se compone únicamente de un abogado con formación en tecnología y programación y un licenciado en sistemas. De cada prueba que se realiza participa un abogado/a experto/a en las tareas de la Subsecretaría con la que se trabaja, quien dedica tiempo extra para estas labores. Para el fortalecimiento institucional del laboratorio se estima necesario separar las tareas de investigación y desarrollo, y dentro de ésta última dividir en áreas dedicadas al diseño de prototipos, desarrollo y seguimiento de prototipos, y

## 2. El Fiscal de Estado y la relevancia institucional de la vista fiscal

El Fiscal de Estado es un Órgano unipersonal creado por la Constitución de la Provincia de Buenos Aires en su art. 155, y cuya principal misión es la defensa del patrimonio del Fisco provincial.

Dicha defensa es ejercida de distintas formas. Una de ellas es mediante la representación de la Provincia – sea actora o demandada – en aquellas causas judiciales en las que se diriman asuntos de índole patrimonial.

Otra de las formas – prevista en los arts. 38 a 40 del Decreto Ley nº 7.543/69 – es mediante su intervención previa a la resolución de cualquier procedimiento administrativo que implique alguna forma de afectación de ese patrimonio, como podría ser una erogación presupuestaria, la disposición de un bien del dominio provincial, o el reconocimiento de un reclamo pecuniario, entre muchos otros ejemplos.

Esta particular atribución se denomina “vista fiscal”. En ella, el Fiscal de Estado, expresa si el procedimiento llevado a cabo – a su entender – se ajusta a las normas jurídicas que resultan aplicables al caso y, puntualmente, si su resolución es acorde – o se contrapone – a los intereses patrimoniales de la Provincia.

Si bien el criterio expresado no es vinculante para la Administración, la gran relevancia institucional de la vista radica en que el Fiscal de Estado tiene la facultad, una vez decidida la cuestión en sentido opuesto a su criterio, de oponerse a la misma y solicitar judicialmente su revocación.

Esto constituye una importante herramienta para fortalecer el control de legalidad y la defensa patrimonial de la Provincia de Buenos Aires que la Constitución le asigna.

## 3. Marco Teórico

---

operaciones, a cuyo efecto resulta imprescindible sumar personal capacitado. Esto además de idear instrumentos que reconozcan la labor del experto/a y permitirle dedicar tiempo útil a su tarea. En esa dirección se ha propuesto la realización de acuerdos con las Facultades de Informática e Ingeniería de la UNLP para sumar pasantes, como asimismo incorporar profesionales formados en informática y con interés en inteligencia artificial, todo ello a fin de poder avanzar con las pruebas que se realizan y proyectar su escala dentro de la organización. Adicionalmente se considera estratégico contar con especialistas que trabajan en los Centros de Inteligencia Artificial de las Universidades Nacionales para iniciar una nueva etapa que complemente las pruebas de concepto de bajo costo que hoy se realizan, y avanzar hacia la adopción de sistemas más sofisticados. A ese efecto se ha identificado y propuesto utilizar el financiamiento disponible de por parte del estado nacional provenientes de la Secretaría de Asuntos Estratégicos y del programa “ImpaCT.AR Ciencia y Tecnología”, como también participar de la convocatoria FITBA propiciada por la Subsecretaría de Ciencia de la provincia.

### 3.1. Automatizar para transformar

Mientras convivamos con Administraciones Públicas que combinan formatos 1.0, 2.0, 3.0<sup>4</sup> y teniendo en miras la necesidad de dejar de hacer “lo mismo” pero con más tecnología, se postula la necesidad de repensar la forma de organizar al Estado desde otro enfoque [4].

Se promueve en ese sentido analizar la actividad administrativa desde un enfoque por tareas, a fin de determinar: a) Las tareas automatizables: presentan un bajo nivel de complejidad, y se destacan por ser actividades rutinarias y repetitivas que demandan metodicidad y practicidad; b) Las semi-automatizables: aquellas que presentan una porción estandarizada y mecánica de tareas, las cuales son potencialmente automatizables, mientras el resto no debido a que implican actividades complejas; c) Las no automatizables: consisten en tareas que requieren de habilidades complejas, como el pensamiento crítico y la resolución de problemas en escenarios de alta incertidumbre, actividades que por ahora no pueden ser reemplazadas [5].

Se afirma que “*Si queremos desburocratizar y humanizar las organizaciones públicas a través de técnicas de inteligencia artificial, es clave que ciertos grupos de tareas se arbolean bajo este enfoque. Por eso, transformar la ‘burocracia imprenta’ o digital en una burocracia inteligente, viene dado por reducir o eliminar pequeñas decisiones que se reflejan en copiar y pegar textos, números, y abrir ventanas digitales para dar respuestas estandarizadas o soluciones simples que se piensan una vez y luego, se repiten cientos o miles de veces*” [5].

Un árbol de decisión, “*representa toda la secuencia de decisiones que nos conducen a la creación del documento*”. Para eso, como primer paso debe identificarse el proceso que se pretende arbolar. Luego están las decisiones convenientes, que son aquellas estrictamente necesarias para llegar al objetivo, entre las que se diferencian: las decisiones de base que determinan el modelo, y las decisiones de color que modifican los datos variables. A su vez dentro de las decisiones de color se encuentran las normativas que son aquellas que la propia norma obliga a tomar, que a su turno se clasifican en: a) decisiones con opciones

---

<sup>4</sup> “...El 1.0 correspondiente a la etapa clásica decimonónica, compuesto por papel, la imprenta y el uso de la máquina de escribir. La Administración 2.0 incorpora al ordenador, los procesadores de texto, la impresora y luego el fax. A partir de Internet, los portales digitales, las aplicaciones móviles y las redes sociales, el sector público en este siglo XXI ha comenzado a migrar hacia el formato 3.0, que se caracteriza por varios conceptos o principios (...) que sólo ciertas tecnologías de la información y de la comunicación pueden lograr: remoto, simple, automático, ágil e instantáneo. Una Administración Pública gestionada en un esquema 3.0, implica transformar los binomios tiempo/espacio-burocracia/ciudadanía a partir de potenciar y actualizar tres supraprincipios rectores: optimización, simplificación y facilitación. En definitiva, hoy en día conviven tres formatos - 1.0, 2.0 y 3.0- dentro de las organizaciones públicas, que a gran escala muestran un lento y complejo camino en el que se ha intentado simplificar y optimizar la gestión de los datos y la información...” Corvalán, Juan G. “Hacia una Administración Pública 4.0: digital y basada en inteligencia artificial. Decreto de Tramitación digital completa”, LA LEY 17/08/2018 , 1. Cita: TR LALEY AR/DOC/1683/2018.

determinadas, denominadas decisiones opcionales, b) decisiones con opciones indeterminadas, llamadas decisiones de contenido, y c) decisiones que dan lugar a otras decisiones y que a diferencia del resto dependen de una decisión previa, identificadas como decisiones de repregunta” [6].

La técnica de confeccionar árboles de decisión se asimila a lo que se conoce como diagrama de secuencia<sup>5</sup> o también denominado flujo de trabajo<sup>6</sup>, teniendo como característica común la finalidad de representar posibles resultados derivados de una serie de decisiones relacionadas. Incluso pueden combinarse, pues en muchos casos confeccionar los árboles de decisión con símbolos de diagramas de flujo puede facilitar su comprensión<sup>7</sup>.

En nuestro caso la confección de árboles de decisión nos permite diseñar un algoritmo que aplicando programación tradicional conduzca a la solución buscada. Lo que es muy diferente a un *aprendizaje basado en árboles de decisión*<sup>8</sup>, donde el concepto es aplicado para modelar modelos predictivos automatizados. En este caso, los nodos representan datos en lugar de decisiones, donde cada uno incluye atributos o reglas de clasificación asociadas a una etiqueta que se ubica al final de la ramificación.

### 3.2. Inteligencia artificial como complemento

Si bien el modelado a través de árboles de decisión nos permite secuenciar un proceso de trabajo aplicando el enfoque de tareas, eso por sí solo no es suficiente para modelar una solución con cierta estabilidad en el tiempo, dado las modificaciones que naturalmente tienen los documentos que sirven de soporte a las decisiones.

---

<sup>5</sup> “El diagrama de secuencia es un tipo de diagrama usado para modelar interacción entre objetos en un sistema según UML. En inglés se pueden encontrar como "sequence diagram", "event-trace diagrams" [https://es.wikipedia.org/wiki/Diagrama\\_de\\_secuencia](https://es.wikipedia.org/wiki/Diagrama_de_secuencia)

<sup>6</sup> “El flujo de trabajo (workflow en inglés) es el estudio de los aspectos operacionales de una actividad de trabajo: cómo se estructuran las tareas, cómo se realizan, cuál es su orden correlativo, cómo se sincronizan, cómo fluye la información que soporta las tareas y cómo se le hace seguimiento al cumplimiento de las tareas. Generalmente los problemas de flujo de trabajo se modelan con redes de Petri” [https://es.wikipedia.org/wiki/Flujo\\_de\\_trabajo](https://es.wikipedia.org/wiki/Flujo_de_trabajo)

<sup>7</sup> Existen varias aplicaciones disponibles en Internet que ofrecen plataformas para asistir al usuario en la confección de árboles de decisión.

<sup>8</sup> “Aprendizaje basado en árboles de decisión es un método comúnmente utilizado en la minería de datos. El objetivo es crear un modelo que predice el valor de una variable de destino en función de diversas variables de entrada.”

[https://es.wikipedia.org/wiki/Aprendizaje\\_basado\\_en\\_%C3%A1rboles\\_de\\_decisi%C3%B3n#cite\\_note-1](https://es.wikipedia.org/wiki/Aprendizaje_basado_en_%C3%A1rboles_de_decisi%C3%B3n#cite_note-1) (General)

Frente a esto, el aprendizaje de máquinas o aprendizaje automático<sup>9</sup> permite entrenar un modelo para que a partir de ciertos datos extraiga patrones para resolver situaciones no previstas. Existen dos tipos.

El aprendizaje supervisado, que permite asignar una clase a un nuevo elemento en base a un conjunto de categorías previamente establecidas. Se basa en un entrenamiento con ejemplos con la solución conocida (supervisado) para crear modelos que permitan clasificar nuevos casos análogos. En esta configuración la experiencia adquirida tiene por objetivo predecir la información faltante. Aquí podemos pensar el entorno como un maestro que "supervisa" al aprendiz proporcionando información adicional (etiquetas) [7, p. 23].

El aprendizaje no supervisado, donde no hay distinción entre los datos de entrenamiento y de prueba. El aprendiz procesa los datos de entrada con el objetivo de generar un resumen o una versión comprimida de esos datos [7, p. 23]. Las clases o categorías no se conocen de antemano, sino que son determinadas en base a las propias similitudes entre los elementos; es decir, se centran en la “medida de similitud” entre elementos.

### 3.3. Procesamiento del lenguaje natural

Los algoritmos de aprendizaje automático no pueden procesar directamente los documentos de texto en su forma original, ya que la mayoría de ellos esperan vectores de *features*<sup>10</sup> numéricos de tamaño fijo, en lugar de documentos de texto de longitud variable. Por lo tanto, durante el preprocessamiento, los documentos de texto son convertidos a una representación acorde.

Uno de los métodos comúnmente utilizados para la representación de los documentos y ampliamente aplicado en la categorización de textos, es el método de bolsa de palabras (BoW, por sus siglas en inglés).

El concepto de "bolsa de palabras" es mencionado por primera vez por el lingüista Zellig Harris en 1954 [8]. El modelo BoW es una representación simplificada utilizada en el procesamiento del lenguaje natural (NLP, por sus siglas en inglés) y la recuperación de información (IR). En la práctica, se utiliza principalmente como herramienta de generación de *features*. Luego de transformar el texto en una "bolsa de palabras", es posible calcular varias medidas para

---

<sup>9</sup> “El aprendizaje automático o aprendizaje automatizado o aprendizaje de máquinas (del inglés, machine learning) es el subcampo de las ciencias de la computación y una rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan. Se dice que un agente aprende cuando su desempeño mejora con la experiencia y mediante el uso de datos; es decir, cuando la habilidad no estaba presente en su genotipo o rasgos de nacimiento. “En el aprendizaje de máquinas un computador observa datos, construye un modelo basado en esos datos y utiliza ese modelo a la vez como una hipótesis acerca del mundo y una pieza de software que puede resolver problemas”. [https://es.wikipedia.org/wiki/Aprendizaje\\_autom%C3%A1tico](https://es.wikipedia.org/wiki/Aprendizaje_autom%C3%A1tico)

<sup>10</sup> “Es una propiedad individual medible o característica de un fenómeno que se observa”. Disponible: [https://en.wikipedia.org/wiki/Feature\\_\(machine\\_learning\)](https://en.wikipedia.org/wiki/Feature_(machine_learning))

caracterizar el texto. El tipo más común de *features* a partir del modelo de bolsa de palabras es la frecuencia de los términos, es decir, la cantidad de veces que aparece un término en el texto [9]. Asimismo permite un modelado basado en diccionario, donde cada documento parece una bolsa (no se considera el orden), que contiene algunas palabras del diccionario [10, p. 39].

La representación de bolsa de palabras es fácil de generar, pero está lejos de ser perfecta [11]. Las frecuencias de términos no son necesariamente la mejor representación de un documento. Palabras comunes como "el", "a", "para", casi siempre tienen la frecuencia de término más alta en el texto, por lo que tener un recuento alto no significa necesariamente que la palabra correspondiente sea más importante.

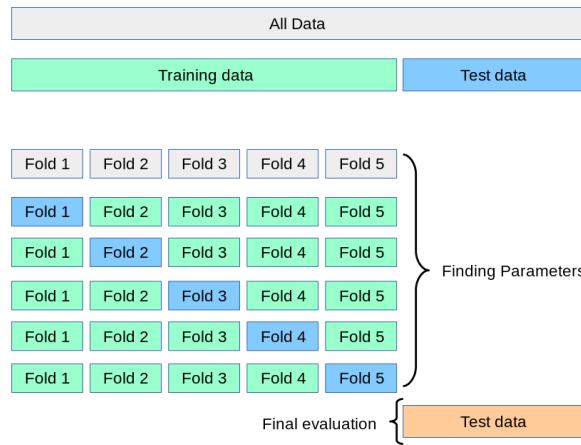
Para abordar este problema, una de las formas populares de "normalizar" las frecuencias del término es ponderar un término por el *inverso de la frecuencia del documento* [9] (Tf-Idf, del inglés Term frequency – Inverse document frequency). Se trata de una medida numérica que expresa cuán relevante es una palabra para un documento de una colección y consiste en un simple giro en el enfoque de la bolsa de palabras. En lugar de ver los recuentos en bruto de cada palabra de cada documento en un dataset, tf-idf mira un recuento normalizado donde cada recuento de palabras se divide por el número de documentos en que aparece esta palabra [11]

Por otra parte, existen diversas técnicas para validar un modelo, comúnmente se utiliza la técnica de validación cruzada (cross-validation)<sup>11</sup>, dividiendo el set de entrenamiento en dos partes (30% testing, 70% training) y entrenando el clasificador con la porción más larga, mientras que la porción más pequeña (validation set) es utilizada para realizar el test de validación. Este método funciona muy bien, pero presenta la desventaja de que el modelo no es entrenado y validado con todos los casos del dataset.

Una variante más sofisticada es el método conocido como *k-fold cross-validation*, a partir de cuya aplicación se divide el dataset en k partes (folds), el proceso se repite k veces (para cada fold), utilizando un subconjunto distinto de datos para validar el modelo entrenado. Luego se promedia el puntaje medido para cada fold, para obtener una estimación más precisa del rendimiento del modelo.

---

<sup>11</sup> La validación cruzada (cross-validation) es un procedimiento de remuestreo que se utiliza para evaluar modelos de aprendizaje automático en una muestra de datos limitada.

**Figura 1.** Cross-validationFuente: [https://scikit-learn.org/stable/modules/cross\\_validation.html#cross-validation](https://scikit-learn.org/stable/modules/cross_validation.html#cross-validation)

## 4. Contexto

### 4.1. Ámbito de trabajo

El ámbito de trabajo para el desarrollo de esta PoC es la Subsecretaría Administrativa y Contenciosa (en adelante SAC). A cada abogado y abogada que cumple tareas en esa área se le asigna un expediente que ingresa por GDEBA<sup>12</sup> para su análisis y proyección de la vista a emitirse, la cual luego de las revisiones por las instancias superiores, es suscripta por el Fiscal de Estado o el funcionario a quien ha delegado esa competencia (Subsecretario o Fiscal Adjunto según el caso).

La implementación de GDEBA a nivel provincial<sup>13</sup> y en el ámbito de la Fiscalía de Estado en particular<sup>14</sup> (ver Imagen 1), permite contar con la digitalización completa del trámite, que la información esté disponible 7x24 para múltiples usuarios, acceder a estadísticas certeras en cuanto al número de las actuaciones en las que se interviene, como asimismo tener a disposición un conjunto de datos para su abordaje mediante técnicas de IA (“small data”).

<sup>12</sup> Sistema de Gestión Documental Buenos Aires. [Online] <https://gdeba.gba.gob.ar/> [Accedido: 12-May-2022].

<sup>13</sup> Por el artículo 2º del Decreto N° 1018/06 se aprobó la implementación del Sistema de Gestión Documental Electrónica Buenos Aires “GDEBA”, como sistema integrado de caratulación, numeración, seguimiento y registración de movimientos de todas las actuaciones de la Administración Pública de la provincia de Buenos Aires (disponible en <https://normas.gba.gob.ar/documentos/0nzA6SM0.html>)

<sup>14</sup> RESO-2017-1-GDEBA-FED del 25/8/2017.

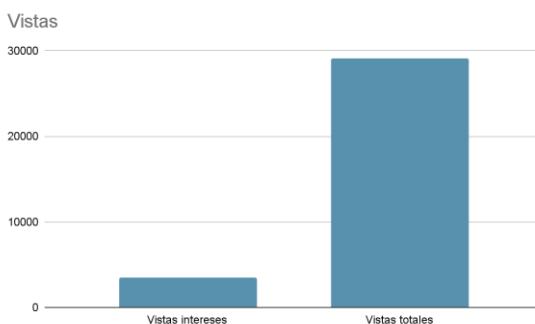


**Imagen 2.** Escritorio GDEBA de un agente de la Fiscalía de Estado.

Con esa ventaja, del cúmulo de trabajo y diversidad de temas tratados en esta área se puso el foco en aquellas que resuelven los reclamos interpuestos por proveedores o contratistas del estado provincial para recibir el pago de intereses por la mora en que incurre el estado en la cancelación de las facturas presentadas al cobro, en el marco de relaciones contractuales de obras o servicios.

Con la participación activa de los expertos que se encargan de confeccionar estas vistas, el primer paso fue verificar la incidencia de los reclamos por intereses sobre el total de temas tratados.

Así, detectamos que entre el 1º de enero de 2021 y el 6 de junio de 2022 se habían despachado 3510 vistas de intereses sobre un total de 29.062 vistas totales, lo que significa que el 12,07% de los expedientes corresponden a este tipo de proceso.



**Tabla 1.** Vistas de intereses sobre vistas totales

#### 4.2. Reseña del circuito administrativo

Cabe destacar que todo procedimiento de reconocimiento y pago de intereses moratorios previstos en el artículo 23 del Reglamento de Contrataciones (Texto Ordenado Decreto N° 3300/72) y en los artículos 45 y 46 de la Ley N° 6021, deben ajustarse a las pautas establecidas por la Resolución Conjunta de la Contaduría General de la Provincia N° 974/09 y la Tesorería General de la Provincia N° 46/09, amén de las disposiciones previstas en el Decreto-Ley N° 7647/70 (Ley de Procedimiento Administrativo).

En particular, para el caso de los reclamos relativos a los intereses establecidos en el referido artículo 23 del Reglamento de Contrataciones, correspondientes a facturas cuya cancelación se hubiere efectuado a través de la Tesorería General de la Provincia, el procedimiento administrativo se puede resumir en los siguientes pasos.

El expediente se inicia con la presentación del proveedor ante la Administración para formalizar su reclamo. Ese primer paso consta de una presentación en papel,

firmada de manera ológrafo, que al momento de ser recibida por la Mesa General de Entradas se le estampa un sello de goma con la fecha de recepción.

El Organismo Contratante, a través de su Dirección de Contabilidad o dependencia equivalente, inicia el trámite en GDEBA e incorpora el reclamo digitalizado, la documentación relativa a la contratación (orden de compra, contrato de locación de inmueble o similar), conjuntamente con la factura reclamada. Asimismo agrega la orden de pago, en la que consta la fecha de vencimiento correspondiente.

La Tesorería General acompaña como archivo embebido el libramiento de pago, en el que se consigna la fecha en la que se canceló la obligación.

A continuación se elevan las actuaciones para la intervención sucesiva de Asesoría General de Gobierno, Contaduría General de la Provincia y Fiscalía de Estado.

Luego vuelven a la repartición de origen y si el reclamo no resulta procedente, se dicta la resolución denegatoria. En caso contrario, de no mediar observación de los organismos de asesoramiento y control, en caso de haberse determinado la viabilidad preliminar del reclamo, se practica la liquidación de intereses y se continúa con el circuito de pago pertinente.

Cuando se trata de la liquidación y pago de intereses por parte de Entidades Descentralizadas y Tesorerías Sectoriales el circuito es similar, con la salvedad que son estas dependencias las encargadas de incorporar los libramientos de pago correspondientes.

#### **4.3. Enfoque por tareas**

Constatada la incidencia de estas vistas en relación a la totalidad de las materias tratadas, siguiendo con la línea de trabajo inaugurada en la PoC aplicada a los amparos por mora y con dicha base teórica [1], volvimos a realizar un análisis de la actividad administrativa desplegada bajo el enfoque de tareas, a fin de determinar si las tareas involucradas eran automatizables, semi-automatizables, o no automatizables<sup>15</sup> [5].

Sobre esa base se concluyó que en efecto presentaba un bajo nivel de complejidad, involucra actividades rutinarias y repetitivas, y por ende eran susceptibles de ser automatizadas con intervención humana reducida.

Luego de relevar las tareas implicadas en la generación de vistas en este tipo de reclamos, se confeccionó un árbol de decisión que nos permitiera diagramar todas las situaciones posibles, identificando los modelos de escrito utilizados para cada caso.

---

<sup>15</sup> Las automatizables presentan un bajo nivel de complejidad, y se destacan por ser actividades rutinarias y repetitivas que demandan metodicidad y practicidad; las semi-automatizables cuentan con una porción estandarizada y mecánica de tareas potencialmente automatizables; y las no automatizables requieren de habilidades complejas, como el pensamiento crítico y la resolución de problemas en escenarios de alta incertidumbre, actividades que por ahora no pueden ser reemplazadas (Corvalán Palumbo, 2019).

En pos de lograr su automatización se identificaron los datos variables de cada modelo, se les asignó un identificador y se procedió a señalar la ubicación del dato en un expediente GDEBA.

A continuación, se elaboró una guía de preguntas y respuestas para asistir al usuario en el camino hacia el tipo de vista pretendido. En caso de ser posible responder las preguntas con los datos disponibles, una a una se van descartando y sólo se realizan aquellas que resultan imprescindibles para completar el total de datos variables requeridos en el caso específico.

## 5. VELOX aplicado a la emisión de la vista en los reclamos por intereses: La solución diseñada.

### 5.1. Extracción de datos, cálculos y análisis del caso.

Con el marco referenciado se avanzó en la solución tecnológica que VELOX podía brindar al caso.

Como hemos dicho antes, la labor del abogado o abogada del área comienza con la asignación del expediente a través de la plataforma GDEBA. En ese momento, se debe hacer uso de la opción de descargar el expediente completo, lo que permite contar con la totalidad de las actuaciones producidas en el expediente bajo un archivo en formato .zip.

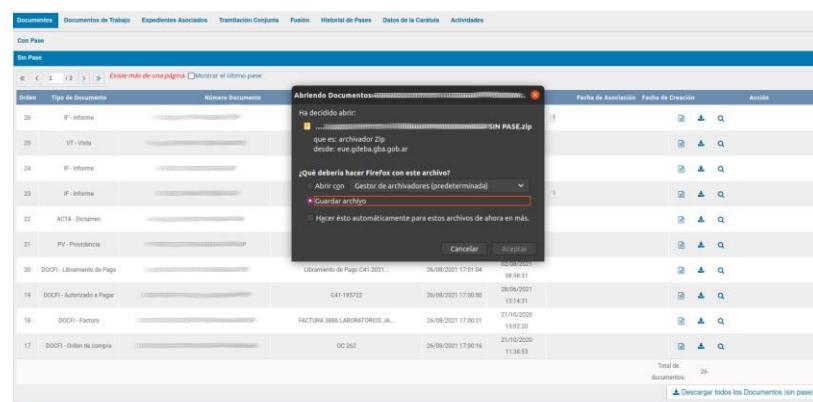
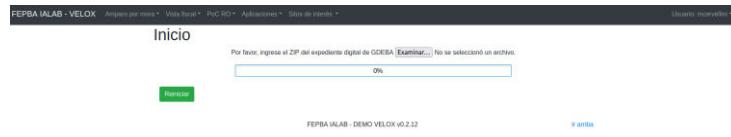


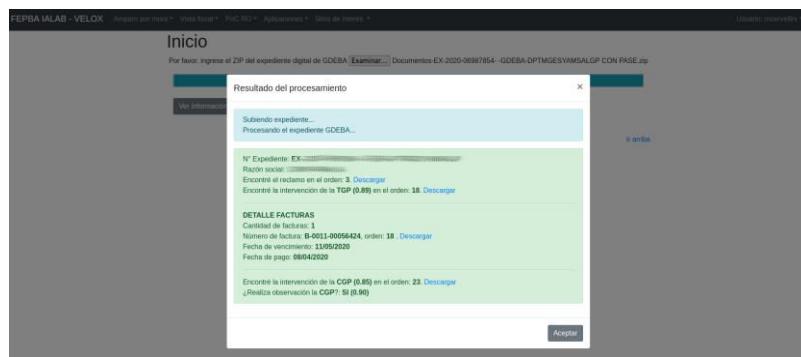
Imagen 3. Descarga del expediente en formato .zip desde GDEBA.

A partir de allí puede iniciarse la asistencia de VELOX, para lo que resulta necesario que el abogado o abogada suba al sistema el expediente descargado de la plataforma GDEBA. Ello activa el proceso de búsqueda de datos guiada por la automatización del proceso, con la interacción con el usuario a través de la interfaz instrumentada en base a las preguntas y respuestas, y la intervención del usuario en caso de ser necesario.



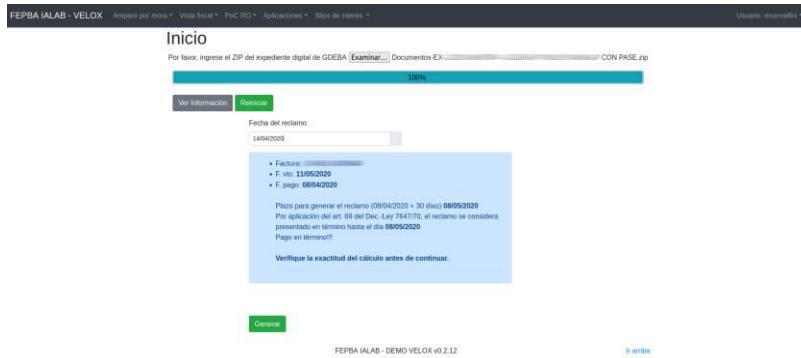
**Imagen 4.** Pantalla de inicio de VELOX para generar vistas.

Una vez identificados los datos disponibles en los documentos que integran el expediente para ser incorporados a la vista a generarse, el sistema está en condiciones de incorporarlos a los modelos y/o de calcular mediante operaciones aritméticas nuevos datos necesarios.



**Imagen 5.** Pantalla de VELOX que muestra los resultados de la búsqueda de documentos dentro del expediente.

A partir del cruce de la información recabada se constata si la factura ha sido pagada en término, si el reclamo es extemporáneo o si corresponde el pago de intereses por mora, luego de lo cual se genera la vista automáticamente.



**Imagen 6.** Pantalla final de VEOX con las conclusiones del trámite y el botón para generar la vista.

## 5.2. La detección de los documentos a partir del uso de expresiones regulares

Para identificar los documentos y extraer la información necesaria se comenzó a trabajar con la premisa de detectar su ubicación (orden) de cada archivo PDF que compone el expediente GDEBA -aplicando técnicas de programación tradicionales- utilizando expresiones regulares<sup>16</sup>.

En base a los distintos casos posibles de reclamos de intereses, se efectuaron pruebas sobre 70 expedientes, a razón de 10 expedientes por cada modelo de vista, consignándose la efectividad de la aplicación para cada uno, y detallándose los inconvenientes detectados.

Asimismo, en base a una muestra de 470 casos estudiados, se efectuó una ponderación de la incidencia de cada caso sobre el total de los reclamos de intereses ingresados<sup>17</sup>.

De esa ponderación surge que la aplicación registra un porcentaje muy alto de efectividad del orden del 95% en los casos cuyos modelos representan un 82% del total. En este circuito los libramientos de pago se encuentran embebidos en la intervención de dicho organismo.

Sin embargo, para otros casos correspondientes a reclamos de uno o más expedientes tramitados por el circuito de la Repartición o reclamos rechazados por extemporáneos o por pago a término, que representan aproximadamente un 12% de los reclamos ingresados, la aplicación no detecta los libramientos de pago, por lo que no genera vista. Esto se debe a que los libramientos no se encuentran embebidos en la intervención de Tesorería General de la Provincia (TGP), sino dispersos en los distintos órdenes del expediente.

Finalmente en situaciones de admisión parcial donde se combinan reclamos válidos con rechazos por reclamo extemporáneo o pago a término, y ascienden a un 6% de los casos ingresados, la aplicación no genera vista correcta, por diversos motivos sobre los que se está trabajando.

## 5.3. Machine learning aplicado a la detección inteligente

---

<sup>16</sup> “En cómputo teórico y teoría de lenguajes formales, una expresión regular, o expresión racional, también son conocidas como regex o regexp, por su contracción de las palabras inglesas regular expression, es una secuencia de caracteres que conforma un patrón de búsqueda. Se utilizan principalmente para la búsqueda de patrones de cadenas de caracteres u operaciones de sustituciones. Las expresiones regulares son patrones utilizados para encontrar una determinada combinación de caracteres dentro de una cadena de texto. Las expresiones regulares proporcionan una manera muy flexible de buscar o reconocer cadenas de texto. Por ejemplo, el grupo formado por las cadenas Handel, Händel y Haendel se describe con el patrón "H(a|ä|ae)ndel".” [https://es.wikipedia.org/wiki/Expresi%C3%B3n\\_regular](https://es.wikipedia.org/wiki/Expresi%C3%B3n_regular)

<sup>17</sup> Cabe señalar que la referida muestra está conformada por los casos ingresados en un período de 12 meses (04/21 – 03/22) al buzón personal de entrada, entendiéndose que puede considerarse representativa de la totalidad de los reclamos ingresados al buzón general.

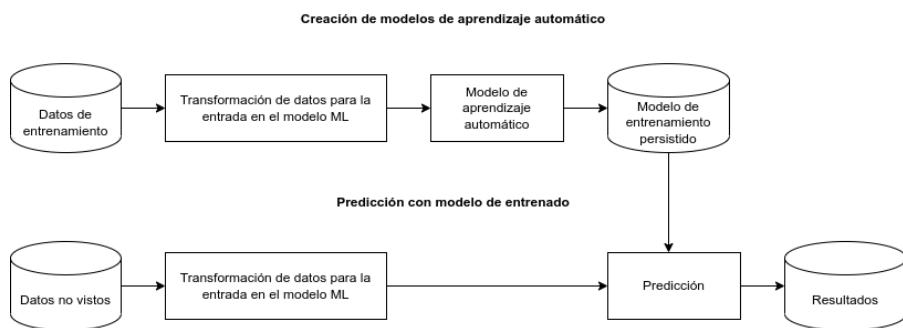
No obstante el buen rendimiento de la solución aplicada al caso, con el correr del tiempo se comenzó a detectar que algunas actuaciones comenzaban a tener variaciones en cuanto a su estructura y contenido, lo que nos llevó a explorar otras posibilidades para el reconocimiento de los documentos.

Sobre la base de la tecnología aplicada para la detección del contenido de las cédulas en el caso de los amparos por mora, se entrenó un nuevo modelo predictivo para la detección de las intervenciones consideradas relevantes para la extracción de información. En particular se buscaba que el modelo detecte de manera inteligente si el PDF correspondía a alguna de las siguientes clases:

- Presentación del reclamo: reclamo presentado por el proveedor o contratista.
- Libramiento de pago: documento con el detalle del pago.
- Informe de TGP: informe presentado por la Tesorería General de la Provincia.
- Informe de CGP: informe presentado por la Contaduría General de la Provincia.

Luego de un exhaustivo relevamiento y selección de documentos extraídos de expedientes correspondientes a cada tipo de situación -según la incidencia de cada uno sobre el total de casos-, se confeccionaron corpus por cada categoría, se entrenó al modelo y se lo testeó para analizar su precisión con la ayuda de los expertos en la materia.

En la figura 2 se puede observar el flujo de trabajo utilizado para la creación de los modelos predictivos y su posterior utilización, a cuyo efecto se aplicó la técnica de aprendizaje automático supervisado, mientras que para la representación de los documentos su utilizó el método de bolsa de palabras (BoW, por sus siglas en inglés), ampliamente aplicado en la categorización de textos.



**Figura 2.** Flujo de trabajo de aprendizaje automático

Las pruebas realizadas sobre un corpus de 1592 intervenciones para entrenamiento y 648 para testeо, arrojó una tasa de acierto muy alta del 98% (Tabla 1).

Enseguida se presentó la duda si no estamos ante un caso de *overfitting*<sup>18</sup>, que precisamente es la situación en la que la predicción del modelo acierta (casi) exactamente con los datos de entrenamiento debido a que se lo ha entrenado con casos muy similares. Esto representa un riesgo pues si a futuro la intervención que queremos detectar deja de respetar el formato actual muy probablemente no pueda identificarla o las tasas de acierto caigan abruptamente.

En nuestro caso constatamos que no se trata de un defecto en la confección del dataset<sup>19</sup>, sino que el alto rendimiento se explica en que los documentos generados por las oficinas intervenientes tienen una estructura que se replica y se mantiene invariable.

	precision	recall	f1-score	support
<b>Reclamo</b>	0.99	0.99	0.99	151
<b>Libramiento</b>	0.99	1.00	0.99	143
<b>Intervención TGP</b>	1.00	0.99	0.99	209
<b>Intervención CGP</b>	1.00	1.00	1.00	145
<b>accuracy</b>			0.99	648
<b>macro avg</b>	0.99	0.99	0.99	648
<b>weighted avg</b>	0.99	0.99	0.99	648

**Tabla 2.** Predictivo para detección inteligente del tipo de intervención. Representación utilizando método BoW y clasificador SGDClassifier

<sup>18</sup> “En aprendizaje automático, el sobreajuste (también es frecuente emplear el término en inglés *overfitting*) es el efecto de sobreentrenar un algoritmo de aprendizaje con unos ciertos datos para los que se conoce el resultado deseado. El algoritmo de aprendizaje debe alcanzar un estado en el que será capaz de predecir el resultado en otros casos a partir de lo aprendido con los datos de entrenamiento, generalizando para poder resolver situaciones distintas a las acaecidas durante el entrenamiento. Sin embargo, cuando un sistema se entrena demasiado (se sobreentrena) o se entrena con datos extraños, el algoritmo de aprendizaje puede quedar ajustado a unas características muy específicas de los datos de entrenamiento que no tienen relación causal con la función objetivo. Durante la fase de sobreajuste el éxito al responder las muestras de entrenamiento sigue incrementándose mientras que su actuación con muestras nuevas va empeorando. En otras palabras, el modelo recuerda una gran cantidad de ejemplos en lugar de aprender a notar características.” <https://es.wikipedia.org/wiki/Sobreajuste>

<sup>19</sup> “Un conjunto de datos (conocido también por el anglicismo dataset, comúnmente utilizado en algunos países hispanohablantes) es una colección de datos habitualmente tabulada.” [https://es.wikipedia.org/wiki/Conjunto\\_de\\_datos](https://es.wikipedia.org/wiki/Conjunto_de_datos)

Finalmente, se realizó una prueba previa al despliegue en producción, para analizar la respuesta del predictivo con casos reales. Para esto se utilizaron una serie de expedientes administrativos por los que tramitan los reclamos por pago de intereses. Se utilizaron un total de once expedientes completos cuyos documentos nunca fueron utilizados en ninguna de las pruebas realizadas hasta el momento, con los resultados que se exponen a continuación.

Predijo correctamente el 100% de los informes producidos por la Tesorería General de la Provincia y de la Contaduría General de la Provincia.

Para estas categorías –dado que se parte de la premisa de detectar una única intervención posible válida– se atendió exclusivamente al porcentaje de probabilidad más alto arrojado por el modelo para la categoría en cuestión, oscilando en todos los casos entre el 93% y el 95%.

Con relación a los Libramientos de Pago de las facturas reclamadas tuvo un porcentaje de aciertos del 90,91%. En este caso, dado que un mismo expediente pueden encontrarse adjuntos varios de estos documentos, se prestó especial atención a las predicciones con mayor probabilidad asignada, advirtiéndose que en todos los casos en que superó el 85% se trataba efectivamente de Libramientos de Pago, sin excepción.

Es dable mencionar que en el único caso en el cual el modelo predijo erróneamente como “Libramiento de Pago” un escrito distinto a los utilizados como corpus de entrenamiento coincidió con la falta de agregación al expediente – por parte de la Repartición actuante – de los Libramientos de Pago pertinentes, asignándole de todas maneras a dicho documento una probabilidad de acierto del 54% solamente.

Finalmente, la clase correspondiente a los reclamos efectuados por las firmas o particulares solicitantes fue la salida que más incertidumbre presentó en las pruebas realizadas, acertando correctamente sólo un 54,54% de predicciones con una probabilidad promedio de 86%.

Se procedió por tanto al análisis particularizado de cada documento al cual el modelo asignó una probabilidad mayor para la categoría 1 que la correspondiente al verdadero Reclamo, destacándose que de los doce (12) casos identificados como “falsos positivos” 8 de ellos (un 66%) corresponden a órdenes de compra, mientras que el porcentaje restante se relaciona con notas o escritos particulares presentados por la firma, diferentes al reclamo.

De ello se desprende la necesidad de profundizar el estudio de estos documentos y de los que integran el corpus de entrenamiento a fin de mejorar el rendimiento del modelo, ajuste que se propone para una etapa futura, donde se deberá trabajar en la optimización del proceso de extracción de la información, cuando se trate de documentos escaneados.

De todas maneras habrá que estar atentos a esta cuestión a futuro e ir reentrenando al modelo si el estado de situación se modifica.

#### **5.4. Resguardo de los datos**

Teniendo en consideración la importancia de dar un uso adecuado de los datos utilizados, así como la necesidad de adoptar los recaudos necesarios para cumplir

con la normativa vigente, se decidió que el prototipo desarrollado no conserve los datos en el servidor. Esto significa que los datos del expediente extraídos para la generación de las vistas, se guardan temporalmente en una carpeta llamada “tmp”, y luego su contenido es eliminado automáticamente mediante una tarea programada en un cron<sup>20</sup>, pues dejan de ser necesarios. Con respecto al modelo predictivo -que es un archivo binario- hay que mencionar que los datasets de entrenamiento se recopilan y trabajan de manera local y luego de realizada dicha tarea se los sube al servidor.

No obstante se está trabajando en un "pipeline"<sup>21</sup> para la captura y almacenamiento de la información que contemple la anonimización de los datos desde su obtención, evitando luego trabajar con datos personales o información sensible, siempre y cuando esto sea posible y la información no sea requerida para entrenar el modelo predictivo.

### 5.5. Conclusión preliminar

Al inicio de las pruebas se le requirió a los expertos del área que cronometren el tiempo que tardaban en despachar un expediente de intereses y promediaron los 5 minutos (siempre considerando que no tengan ninguna interrupción o distracción).

Con el uso de VELOX solo resulta necesario que el usuario ingrese los datos para un campo faltante (fecha de inicio del reclamo<sup>22</sup>) para estar en condiciones de generar automáticamente la vista fiscal, lo que insume no más de 5 segundos con un ahorro de tiempo estimado del orden del 98.33%.

Estos tiempos, en la actualidad, no varían según la técnica que esté detrás. Es que si bien utilizando expresiones regulares se obtuvieron resultados aceptables, lo cierto es que a futuro es dable esperar un mejor rendimiento y efectividad aplicando técnicas de machine learning, pues resulta natural que con el paso del tiempo los expedientes comiencen a tener ciertas variaciones en lo que respecta tanto a la ubicación de los documentos (orden) como el contenido mismo de cada intervención (aunque en este último caso habrá que estar atentos a si los cambios afectan la predicción en función de lo mencionado en el Punto 5.3).

## 6. Avances en progreso

Se probará la técnica de reconocimiento de entidades nombradas (NER por sus siglas en inglés) utilizando la librería spaCy<sup>23</sup>, para detectar la razón social o el

---

<sup>20</sup> Es un administrador de procesos en segundo plano (demonio) que ejecuta procesos o scripts a intervalos regulares (por ejemplo, cada minuto, día, semana o mes).

<sup>21</sup> Cadena de procesos conectados de forma tal que la salida de cada elemento de la cadena es la entrada del próximo.

<sup>22</sup> Este dato surge del sello de goma que se estampa en el reclamo en formato papel que presenta el reclamante.

<sup>23</sup> spaCy [Online]. Disponible: <https://spacy.io>. [Accedido: 06-Jun-2022].

nombre del proveedor que reclama el pago de intereses, a partir de un modelo<sup>24</sup> entrenado para este propósito con el listado completo de los proveedores inscriptos en el Registro que funciona en la CGP.

Luego de las pruebas realizadas se pudo observar que el diagrama original de decisión original dejaba abierto una serie de posibilidades irresueltas que ya fueron reseñadas en el punto 4.2 y que era posible reformularlo y ampliarlo de manera de permitir resolver adecuadamente todos los casos, sobre lo que se continúa trabajando.

Respecto a modelo predictivo, se están realizando pruebas para evitar un sobreajuste del modelo, siguiendo las recomendaciones que indican que en estos casos deben utilizarse algunas de las técnicas disponibles, que en nuestro caso están dirigidas a perturbar algunos de los datos de entrenamiento [12, p. 35-36]. Asimismo se está avanzando en la posibilidad de detectar el contenido de los informes de la Contaduría General de la Provincia, para distinguir dos clases posibles: por un lado, está la posibilidad de que el informe formule o no observaciones al trámite, y por el otro, que aconseje el rechazo del reclamo por extemporáneo o por haberse realizado el pago a término.

Finalmente, se avanza firmemente en la interoperabilidad con GDEBA ya que se trata de un aspecto que resultará determinante para el óptimo funcionamiento de VELOX. Desde el Laboratorio se ha suscrito con el Director de Software Público un “Acta de Constitución de Proyecto IOP050 - GDEBA - FEPBA IALAB- VELOX”<sup>25</sup>, y actualmente se encuentra en su fase de desarrollo, habiendo logrado su integración de manera exitosa con la API de GDEBA para ambientes bajos<sup>26</sup>.

## 7. Un cambio de paradigma resulta necesario

El desafío próximo, como se ha repetido infinidad de veces en los últimos tiempos, pasa por repensar las posibilidades que brindan las nuevas tecnologías y dejar atrás la respuesta refleja de replicar digitalmente los procesos y procedimientos concebidos para el soporte papel.

---

<sup>24</sup> spaCy ofrece una opción para agregar clases arbitrarias a los sistemas de reconocimiento de entidades y actualizar el modelo para incluir incluso los nuevos ejemplos además de las entidades ya definidas dentro del modelo.

<sup>25</sup> Dicho convenio fue suscripto a fin de lograr la interoperabilidad que permita optimizar procesos y la generación automática de documentos en el ámbito de Fiscalía de Estado, permitiendo incubar prototipos tecnológicos aplicando soluciones de inteligencia artificial (IA) en un ambiente controlado, accediendo a la información contenida en los expedientes administrativos que tramitan en esa plataforma. Una vez concluido la efectiva integración de los sistemas VELOX y GDEBA permitirá que la aplicación VELOX: • consulte todos los expedientes electrónicos depositados en los buzones de Fiscalía de Estado dentro de la plataforma GDEBA; • acceda a la totalidad de la información contenida en dichos expedientes electrónicos; • cargue de manera automática los requerimientos emitidos por Fiscalía de Estado en los expedientes contenidos dentro de la plataforma GDEBA.

<sup>26</sup> coadyuva a esta evolución que hoy podamos contar con una Infraestructura Virtualizada en funcionamiento donde se aloja VELOX, producto del “Acta de Asistencia Técnica de Nube (Cloud Computing)” suscrita entre el FEPBA IALab y la DPSIT#, con la seguridad y el respaldo que apareja una solución de este tipo, que además es brindada por el propio estado provincial sin gasto para el Organismo.

En este sentido, resulta insoslayable explorar otros caminos que permitan superar las limitaciones que presenta el lenguaje natural para ser entendido por las máquinas, en especial el proveniente de abogados y abogadas<sup>27</sup>.

En otros países ya se han dado algunos pasos en esta dirección, a través del registro como datos estructurados de la información que hasta entonces sólo estaba disponible en lenguaje natural<sup>28</sup>.

Pensemos por ejemplo si en lugar de tener un expediente como el que nos ocupa en este estudio, con numerosas intervenciones plasmadas en otros tantos documentos de los que pretendemos extraer la información aplicando distintas técnicas de procesamiento del lenguaje, existiera una plataforma que permita al reclamante ingresar por campos los datos de su reclamo, y que la Administración pudiera analizarlos registrando las intervenciones de cada organismo competente para resolver la petición con tan sólo emitir un documento que refleje la decisión final del órgano competente, ¿se imaginan el ahorro de tiempos y recursos que significa algo así?<sup>29</sup>.

## 8. Conclusiones generales

Esta investigación, como las anteriores PoC llevadas adelante, siempre están guiadas por una máxima de trabajo que es no aplicar técnicas sofisticadas o complejas donde existen soluciones más simples para resolver el problema.

---

<sup>27</sup> “La investigación realizada por *Eric Martínez, Francis Mollica y Edward Gibson (Poor writing, not specialized concepts, drives processing difficulty in legal language)*, constata que la estructura, términos y gramática utilizados en los documentos legales son, por un lado, muy específicos del sector, y por otro, y en consecuencia, muy alejados de los utilizados en el lenguaje común.” [13]

<sup>28</sup> “Recientemente se ha puesto en marcha una nueva interfaz para la recepción de las reclamaciones de los contribuyentes – “e-Defesa” - con la promesa de registrar como datos estructurados toda la información que hasta ahora sólo estaba disponible en lenguaje natural. e-Defesa facilita la elaboración de una solicitud de apelación tributaria, presentando al contribuyente sugerencias de alegaciones relevantes para cada tipo de infracción incluida en la notificación de la declaración de impuestos, e indicando los documentos que deben entregarse a la RFB. Las infracciones fiscales, junto con las correspondientes contralegaciones, se codifican de manera comprensible, de modo que es necesario un uso mínimo de lenguaje natural y los puntos clave del proceso contencioso se pueden almacenar como datos estructurados.” [14]

<sup>29</sup> En este sentido la Declaración de Santo Domingo aprobada en la XX Conferencia Iberoamericana de Ministras y Ministros de Administración Pública y Reforma del Estado, celebrada el día 25 de mayo de 2022 en su punto 28 alienta a “Acelerar las políticas de digitalización desarrolladas y acordadas en las recientes Reuniones Ministeriales de Administración Pública, replanteándose la existencia y pertinencia de procedimientos administrativos innecesarios en la sociedad actual, así como los procesos a digitalizar que creen efectivamente valor público”. [Online]. Disponible: <https://clad.org/wp-content/uploads/2022/06/Declaracion-SantoDomingo-ES-06-2022.pdf> [Accedido: 17-Jun-2022]

En este sentido, si bien al principio de la prueba no resultó necesario aplicar machine learning para lograr la generación automática de las vistas, con el tiempo se detectó que las variaciones en la ubicación (orden), el formato y contenido de la intervenciones comenzaba a tener variaciones que dificultan hallar los datos variables. La aplicación de las técnicas de machine learning desarrolladas para VELOX nos permitió afrontar esta situación y lograr resultados altamente satisfactorios en lo que refiere a la detección de documentos.

Continúa siendo una asignatura pendiente explorar técnicas más complejas de machine learning, así como lograr avances en la integración e interoperabilidad con sistemas interno y externos del Organismo.

Desde lo institucional, a la reciente creación del “Laboratorio de Inteligencia Artificial para la Fiscalía de Estado” (FEPBA IALab)<sup>30</sup> se le suma la constitución de la Subsecretaría de Innovación y Desarrollo Tecnológico de la Fiscalía de Estado<sup>31</sup>, lo que permitirá avanzar en el despliegue de soluciones de IA en otras áreas del organismo, como asimismo facilitar la interacción con las áreas TI para alcanzar la gobernanza de datos e interoperabilidad deseadas.

## 9. Referencias

1. Cervellini M. C., Carbone M., “VELOX: Inteligencia artificial aplicada al proceso de amparo por mora. Un prototipo para la gestión interna de la Fiscalía de Estado de PBA”. [Online]. Disponible: <https://50jaiio.sadio.org.ar/pdfs/sid/SID-06.pdf>. [Accedido: 06-Jun-2022].
2. Cervellini M. C., Carbone M., “Utilización de un modelo predictivo como soporte para la revisión de los recursos ordinarios de apelación”. [Online]. Disponible: <https://49jaiio.sadio.org.ar/pdfs/sid/SID-03.pdf>. [Accedido: 06-Jun-2022].
3. Grafe F. ¿Es efectiva la defensa jurídica del Estado? Perspectivas sobre la gestión del riesgo fiscal en América Latina, Documento de debate # IDB-DP-338, BID, 2014. [Online]. Disponible: <https://publications.iadb.org/es/publicacion/13725/es-efectiva-la-defensa-juridica-del-estado-perspectivas-sobre-la-gestion-del> [Accedido: 27-Mar-2023].
4. Corvalán Juan Gustavo y Ana Paula Montenegro, “Black Mirror ‘Bandersnatch’ y el derecho administrativo 4.0”. Diario Administrativo Nro. 247 - 13.08.2019. [Online]. Disponible: <https://dpicuantico.com/sitio/wp-content/uploads/2019/08/Doctrina-Black-Mirror-%E2%80%9CBandersnatch.docx.pdf> [Accedido: 27-Mar-2023].
5. Corvalán J. G. y Palumbo J., “Inteligencia Artificial y Trabajo. Explorando un nuevo paradigma laboral”. Diario Laboral Nro. 210 - 11.07.2019. [Online]. Disponible: <https://dpicuantico.com/sitio/wp-content/uploads/2019/07/Doctrina-Corval%C3%A1n-y-Palumbo-11-07-1.pdf> [Accedido: 27-Mar-2023].

<sup>30</sup> Aún en etapa de desarrollo institucional. A este efecto se contó con el apoyo de la Subsecretaría de Empleo Público y Gestión de Bienes de la PBA mediante una asistencia técnica formalizada el 22-9-2021 a través de un “Plan de trabajo para el fortalecimiento institucional” aprobado por firma conjunta, lo que permitió contar con la asistencia de una experta en modernización y elaborar documentos de base.

<sup>31</sup> RESO-2022-57-GDEBA-FED del 11/3/2022.

6. Tanno Natalia, "Cómo automatizar procesos de compras públicas. Experiencia Prometea compras", Diario Administrativo Nro. 255 - 08.10.2019 - DPI Cuántico. [Online]. Disponible: <https://dpicuantico.com/sitio/wp-content/uploads/2019/10/Doctrina-administrativo-08-10.pdf> [Accedido: 27-Mar-2023].
7. Shai Shalev-Shwartz and Shai Ben-David. Understanding Machine Learning: From Theory to Algorithms. 2014. Cambridge University Press. [Online]. Disponible: <http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/understanding-machine-learning-theory-algorithms.pdf> [Accedido: 24-Mar-2019].
8. Harris, Zellig (1954). "Distributional Structure". Word. 10 (2/3): 146–62.
9. Bag-of-words model [Online]. Disponible: [https://ipfs.io/ipfs/QmXoypizjW3WknFjNKLwHCnL72vedxjQkDDP1mXWo6uc0/wiki/Bag\\_of\\_words\\_model.html](https://ipfs.io/ipfs/QmXoypizjW3WknFjNKLwHCnL72vedxjQkDDP1mXWo6uc0/wiki/Bag_of_words_model.html) [Accedido: 20-Mar-2019].
10. Proyecto Final de Carrera: Aplicación del modelo Bag-of-Words al reconocimiento de imágenes. Julio del 2009. Consultado el 11 de noviembre de 2014. [Online]. Disponible: [https://e-archivo.uc3m.es/bitstream/handle/10016/7107/PFC\\_Sara\\_Pardo\\_Fejoo.pdf?sequence=1](https://e-archivo.uc3m.es/bitstream/handle/10016/7107/PFC_Sara_Pardo_Fejoo.pdf?sequence=1) [Accedido: 18-Mar-2019].
11. Amanda Casari, Alice Zheng. Feature Engineering for Machine Learning. O'Reilly. Abril 2018. [Online]. Disponible: <https://www.oreilly.com/library/view/feature-engineering-for/9781491953235/ch04.html> [Accedido: 21-Mar-2019].
12. Sánchez Ávalos, R.; González, F.; Ortiz, T., "Uso responsable de la IA para las políticas públicas: manual de ciencia de datos", BID, 2021.[Online]. Disponible: <https://publications.iadb.org/es/ia-responsable-manual-tecnico-ciclo-de-vida-de-la-inteligencia-artificial> [Accedido: 27-Mar-2023].
13. The Technolawgist. Law: Tech. Future. "La impenetrabilidad de los textos legales y sus vías de solución: ¿por qué no te entiendo abogado?". [Online]. Disponible: <https://www.thetechnolawgist.com/2022/03/17/la-impenetrabilidad-de-los-textos-legales-y-sus-vias-de-solucion-por-que-no-te-entiendo-abogado/> [Accedido: 27-Mar-2023].
14. Ministério da Economía, Receita Federal de Brasil, CIAT, DT-04-2021: Inteligencia Artificial para la Decisión Expedita y el Análisis de Riesgos en el Contencioso Administrativo Tributario Federal Brasileño, CIAT 2021. [Online]. Disponible: <https://www.ciat.org/dt-04-2021-inteligencia-artificial-para-la-decision-expedita-y-el-analisis-de-riesgos-en-el-contencioso-administrativo-tributario-federal-brasileño/> [Accedido: 27-Mar-2023].