

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

Mariano Cervellini¹, Miguel Carbone¹, Blas Szelagowski¹ Federico Gonzalez Oria¹

¹Fiscalía de Estado de la Provincia de Buenos Aires, Av. 1 esq 60, 1342

La Plata (1900), Argentina

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

Abstract

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.

Keywords: Automatización, Machine Learning, Clasificación, Aprendizaje supervisado.

1. Introducción

VELOX [1] es un prototipo que surge de una prueba de concepto (PoC) desarrollada durante el período 2020-2021¹ en el ámbito de la Fiscalía de Estado de

¹ A partir del mes de julio de 2021 se instrumentó una prueba piloto por 6 meses que fue prorrogada y a la fecha aún se encuentra en desarrollo. Esa primera etapa 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

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].

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 (Grafe, 2014), nuestro trabajo se focaliza en aprovechar el rendimiento exponencial de la 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.

En esas premisas, los promisorios resultados obtenidos y la participación activa de los propios agentes en un ámbito institucionalizado², 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.

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 n° 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.

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.

² 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.

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. Contexto

3.1. 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 GDEBA³ 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 provincial⁴ y en el ámbito de la Fiscalía de Estado en particular⁵, 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*”).

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 1° 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,

³ Sistema de Gestión Documental Buenos Aires. [Online] <https://gdeba.gba.gob.ar/>. [Accedido: 12-May-2022].

⁴ Por el artículo 2° 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>)

⁵ RESO-2017-1-GDEBA-FED del 25/8/2017.

lo que significa que el 12,07% de los expedientes corresponden a este tipo de proceso.

3.2. 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 ológrafa, 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.

3.3. 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 (Corvalán Palumbo, 2019).

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.

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.

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

4.1. 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.

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.

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.

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.

4.2. 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.

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 ingresados⁶.

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 porcentaje del 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 representan aproximadamente un 6% de los casos ingresados, la aplicación no genera vista correcta, por diversos motivos sobre los que se está trabajando.

4.3. Machine learning aplicado a la detección inteligente

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.

⁶ 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.

- 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 1 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 se utilizó el método de bolsa de palabras (BoW, por sus siglas en inglés), ampliamente aplicado en la categorización de textos.

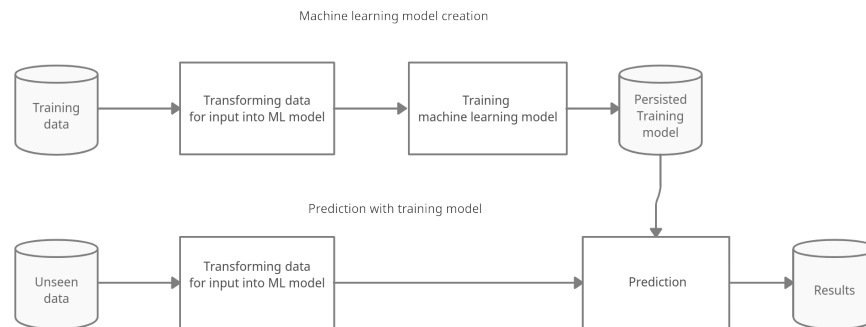


Figura 1. Machine Learning Workflow

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

Enseguida se presentó la duda si no estamos ante un caso de *overfitting*, 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, sino que el alto rendimiento se explica en que los documentos generados por las oficinas intervinientes tienen una estructura que se replica y se mantiene invariable.

Tabla 1. Predictivo para detección inteligente del tipo de intervención. Representación utilizando método BoW y clasificador SGDClassifier

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

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.

4.4. 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 cron⁷, 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"⁸ 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.

4.5. 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 reclamo⁹) 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%.

⁷ 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).

⁸ Cadena de procesos conectados de forma tal que la salida de cada elemento de la cadena es la entrada del próximo.

⁹ Este dato surge del sello de goma que se estampa en el reclamo en formato papel que presenta el reclamante.

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 4.3).

5. Avances en progreso

Se probará la técnica de reconocimiento de entidades nombradas (NER por sus siglas en inglés) utilizando la librería spaCy¹⁰, para detectar la razón social o el nombre del proveedor que reclama el pago de intereses, a partir de un modelo¹¹ 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 [6, 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-

¹⁰ spaCy [Online]. Disponible: <https://spacy.io>. [Accedido: 06-Jun-2022].

¹¹ 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.

VELOX”¹², y actualmente se encuentra en su fase de desarrollo, habiendo logrado su integración de manera exitosa con la API de GDEBA para ambientes bajos¹³.

6. 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.

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 abogadas¹⁴.

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 natural¹⁵.

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

¹² 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.

¹³ 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.

¹⁴ “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.” [9]

¹⁵ “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 contra-alegaciones, 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.” [7]

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?¹⁶.

7. Conclusiones generales

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

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 las 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)¹⁷ se le suma la constitución de la Subsecretaría de Innovación y Desarrollo Tecnológico de la Fiscalía de Estado¹⁸, 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.

8. Referencias

¹⁶ 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]

¹⁷ 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 un experto en modernización y elaborar documentos de base.

¹⁸ RESO-2022-57-GDEBA-FED del 11/3/2022.

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. Corvalán J. G., “Inteligencia Artificial GPT-3, PretorIA y Oráculos Algorítmicos en el Derecho”. Publicado en International Journal of Digital Law. Belo Horizonte. Año 01 . n. 01 . enero/abril 2020.
4. Corvalán J. G. y Palumbo J., “Inteligencia Artificial y Trabajo. Explorando un nuevo paradigma laboral”. Diario Laboral Nro. 210 - 11.07.2019.
5. 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.
6. Longo Martínez F., “La Administración Pública en la era una gobernanza exploratoria”, Revista Vasca de Gestión de Personas y Organizaciones Públicas - Núm. especial 3/2019.
7. Ministério da Economia, 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. 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/>
8. Sánchez Ávalos, R.; González, F.; Ortiz, T., “Uso responsable de la IA para las políticas públicas: manual de ciencia de datos”, BID, 2021.
9. The Technolawgist. Law: Tech. Future. “La impenetrabilidad de los textos legales y sus vías de solución: ¿por qué no te entiendo abogado?”. 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/>