

## Producción de indicadores de funcionamiento de la Corte Suprema mediante el uso de de Inteligencia Artificial y Procesamiento de Lenguaje Natural

Juliana Delle Ville<sup>1</sup>, Diego Martínez<sup>2</sup>, Leandro Antonelli<sup>1,3</sup>, Julián Grigera<sup>1,4,5</sup> y Leandro Giannini<sup>2</sup>

<sup>1</sup> LIFIA, Fac. de Informática, Univ. Nac. de La Plata

<sup>2</sup> Instituto de Derecho Procesal, Univ. Nac. de La Plata

<sup>3</sup> CAETI - Facultad de Tecnología Informática - Universidad Abierta Interamericana

<sup>4</sup> CONICET, Argentina, <sup>5</sup> CICPBA, Argentina

{jdelleville, lanto, juliang}@lifia.info.unlp.edu.ar,  
diego\_gmartinez@hotmail.com, lgiannini@gmail.com

**Abstract.** Al ser el tribunal de última instancia del país y el custodio e intérprete final de la Constitución y de las leyes federales, la Corte Suprema (CS) tiene un papel central en el sistema institucional. Sus fallos sirven para guiar la interpretación y el desarrollo del derecho, para determinar los alcances de los derechos y garantías fundamentales de los ciudadanos, para dirimir conflictos suscitados entre los distintos poderes del gobierno nacional o controversias interjurisdiccionales típicas de un país federal como la Argentina. Sin embargo, la CS no publica estadísticas completas, actualizadas y sistematizadas que permitan a la comunidad jurídica y a la sociedad en general supervisar el funcionamiento del tribunal más importante del país y evaluar el mérito de cualquier propuesta de reforma de su integración, competencia y rol institucional. Hasta ahora, se viene aplicando un método personalizado de lectura, análisis y carga de la información con la que se producen los informes de forma manual, lo que consume mucho tiempo y personal. El objetivo de esta propuesta es agilizar el procesamiento de la información publicada por la CS para poder generar datos estadísticos de calidad. Para esto, se utilizaron herramientas como Inteligencia Artificial (IA) y Procesamiento de Lenguaje Natural (PLN). En este primer trabajo, nos enfocamos en predecir materia, vía de acceso y origen de los fallos procesados. Los primeros resultados muestran niveles de precisión superiores al 90% en la predicción de estos aspectos. Este trabajo se enmarca en una iniciativa conjunta del Instituto de Derecho Procesal y el Centro LIFIA, ambos pertenecientes a la Universidad Nacional de La Plata.

**Keywords:** Procesamiento de lenguaje natural, inteligencia artificial, digitalización judicial, corte suprema, transparencia judicial, acceso a la información

### 1 Introducción

Desde 2016, el Instituto de Derecho Procesal de la Universidad Nacional de La Plata, viene desarrollando un proyecto de investigación sobre el Análisis de la Producción Jurisdiccional de la Corte Suprema de Justicia de la Nación (CS). El objetivo del

proyecto es producir información vacante sobre la actuación jurisdiccional de la CS y promover el acceso a la información sobre dicho desempeño. Dicha línea de investigación ha producido una serie de trabajos en los que se presenta información inédita sobre el funcionamiento del tribunal [1-4], incluyendo datos sobre la cantidad de fallos dictados anualmente, cantidad de causas resueltas a través de dichas decisiones, incidencia de las vías de acceso al tribunal en su agenda jurisdiccional, tasa de admisión y de éxito de las apelaciones que tramitan ante sus estrados, duración de los procesos, cohesión de sus integrantes al decidir (frecuencia de las decisiones unánimes o divididas), entre otros.

Producir esta clase de información es importante, dada la relevancia del rol que desempeña la CS en el diseño de nuestras instituciones. La CS es el máximo tribunal judicial de la Argentina, responsable de interpretar en última instancia la Constitución nacional y las leyes federales, así como de corregir excepcionalmente desvíos graves y evidentes de las decisiones finales de los restantes tribunales superiores del país (por ejemplo cámaras nacionales de apelación y supremas cortes provinciales), a través de la doctrina de la arbitrariedad de sentencias. Al ser el tribunal de última instancia del país y el custodio e intérprete final de la Constitución y de las leyes federales, la CS tiene un papel central en el sistema institucional. Sus fallos sirven para guiar la interpretación y el desarrollo del derecho, para determinar los alcances de los derechos y garantías fundamentales de los ciudadanos (especialmente en los frecuentes casos de colisión entre derechos reconocidos por la constitución misma), para dirimir conflictos suscitados entre los distintos poderes del gobierno nacional (judicial, legislativo y ejecutivo) o controversias interjurisdiccionales típicas de un país federal como la Argentina.

La CS no publica series estadísticas completas, actualizadas y sistematizadas que permitan acceder al tipo de información producida por el Instituto de Derecho Procesal, pese a que se trata de datos fundamentales para: a) analizar y supervisar el funcionamiento del tribunal más importante del país; y b) evaluar el mérito de cualquier propuesta de reforma de su integración, competencia y rol institucional. Esta falencia estadística debilita sensiblemente la calidad de la discusión pública sobre un tema que tiene directo impacto en la vigencia plena del estado de derecho en Argentina.

A fines de 2023, el Instituto de Derecho Procesal y el Centro LIFIA, ambos de la UNLP, decidieron emprender un proyecto conjunto para aplicar herramientas de Inteligencia Artificial (IA) y Procesamiento de Lenguaje Natural (PLN) en el análisis de la producción jurisdiccional de la CS.

El problema central a abordar es la eficiencia del relevamiento. Hasta ahora, el Instituto de Derecho Procesal viene aplicando un método personalizado de lectura, análisis y carga de la información con la que se producen los informes. Ello implica: a) asignar capital humano calificado a tareas que, en muchos casos, podrían ser automatizadas mediante la aplicación de herramientas de IA y PLN; y b) insumir un tiempo considerable en la producción de los informes finales, que terminan siendo conocidos y publicados casi dos años después de la finalización del año judicial examinado.

El objetivo del proyecto colaborativo es agilizar el procesamiento de la información aplicando automatización con la ayuda de IA y PLN, pero sin perder precisión, mejorando el rendimiento de la actividad de los investigadores que hoy

desarrollan esta labor. Esto permitiría diseñar una herramienta de acceso para que la ciudadanía pueda conocer los resultados de los relevamientos producidos. Para ello, se propuso trabajar en herramientas que permitan, entre otras funciones:

- Descargar automatizadamente los documentos y fuentes que contienen la información pertinente (por ejemplo, descarga de fallos y dictámenes y de otra información obrante en sitios de acceso público).
- Extraer de los documentos y demás fuentes relevantes la información, pertinente para la investigación.
- Construir con la información extraída una base de datos.
- Clasificar y procesar los datos extraídos a través de herramientas de IA y PLN, para producir indicadores y métricas sobre la producción jurisdiccional de la CS
- Diseñar una plataforma de acceso a la información que permita acceder de modo ágil y transparente a los indicadores y métricas

Luego de las reuniones de trabajo desarrolladas entre directivos e investigadores del Instituto de Derecho Procesal y del Centro LIFIA de la UNLP, se acordó avanzar progresivamente en el abordaje de dichos objetivos, evaluando en cada fase la eficacia de las herramientas desarrolladas.

En el presente trabajo se exponen los resultados de la primera fase, que consistió en desarrollar un algoritmo que permita la identificación, clasificación y agrupamiento de la producción jurisdiccional de la CS en relación a las siguientes temáticas:

**A.-Materia.** En este caso se utilizan reglas heurísticas las que se complementan con Machine Learning para identificar la materia sobre la que versa cada una de las causas que componen la agenda jurisdiccional de la Corte. Esta información nos permite conocer la composición cualitativa de la actuación del máximo tribunal nacional distinguiendo cuantos asuntos civiles, laborales, contenciosos administrativos, penales, etc. transitan ante sus estrados anualmente.

**B.-Vías de acceso.** A partir de la implementación de reglas heurísticas, el algoritmo clasifica y obtiene la información relevante en relación a cada una de las vías de acceso ante la CS. Se distinguen los tres grupos principales de actuación jurisdiccional de la Corte Suprema: la competencia apelada (que incluye los casos que llegan a la Corte por recursos deducidos contra decisiones de tribunales inferiores), los juicios originarios (en los que la CS entiende como tribunal de primera y única instancia) y los agrupados en la categoría “otros” en lo que se incluyen una gama de variadas de decisiones que no ingresan en las restantes categorías (v.gr. cuestiones de competencia, honorarios, etc.).

**C.-Origen de la causa.** A partir de una adecuada parametrización, el algoritmo clasifica y arroja resultados que permiten distinguir y sistematizar de dónde provienen los casos llevados ante los estrados de la CS. Esto permite obtener datos confiables sobre el aporte de cada jurisdicción en la abultada agenda del tribunal. Se distingue inicialmente según los fallos dictados y las causas resueltas prevengan: de la justicia provincial; o de la justicia nacional.

## 2 Trabajo Relacionado

Esta sección describe trabajos similares en la problemática descripta. Se dividió la sección en dos subsecciones. En primer lugar se describen trabajos similares en donde se usa IA y PLN en la justicia en general. Y luego se describen trabajos que se ocupan de realizar categorización y extracción de información, también utilizando IA y PLN.

### 2.1 Marco general del uso de la IA y PLN en la justicia

Serventich [9] sostiene que el uso de la IA es una buena forma de agilizar el proceso judicial, pero al mismo tiempo, critica que es una caja negra y no es transparente el proceso de pensamiento para poder llegar a un resultado. También aclara que dichos modelos podrían ser sesgados por los datos de entrenamiento y relata un caso en que una IA clasificaba a la persona de color de una forma y a una persona de tez blanca de otra. Según UNESCO [8] hoy en día hay un fuerte uso de aplicaciones de IA con el fin de buscar antecedentes para dar soporte a los argumentos. Además, ciertos bufetes de abogados utilizan técnicas basadas en IA para calcular estadísticas que puedan utilizar para pronosticar resultados de casos con el fin de evaluar la situación y poder asesorar al cliente. Este artículo, propone una estrategia para analizar el circuito que llevaron los casos y poder así conocer el funcionamiento de los componentes judiciales del estado. Según el informe Bustos et al [10], existen cuatro iniciativas de innovación en el ámbito judicial de Argentina entre ellos destaca mecanismos de automatización inteligentes de documentos judiciales y PROMETEA. Pero estas iniciativas tienen una muy limitada escalabilidad y un grado incipiente de desarrollo, salvo PROMETEA que se convirtió en una política definida. Argentina no presenta políticas públicas ni incentivos institucionales para innovar en este campo y la gran mayoría de proyectos se encuentran en CABA, jurisdicción que tiene el mayor PBI per cápita del país. Por lo que es necesario encontrar una forma de fomentar la iniciativa.

### 2.2 Categorización y Extracción de información en Textos Judiciales

Sposito et al. [11] plantean una extracción mediante expresiones regulares para reconocer entidades nombradas clasificarlas y reemplazarlas. Si bien es una estrategia interesante, nuestro trabajo utiliza reglas que son complementadas con algoritmos de IA, por lo cual, la identificación y extracción es más rica. También hay trabajos que buscan mediante procesamiento de lenguaje natural identificar indicios de narcotráfico en mensajes de Whatsapp (Cassanelli[12]). Esta propuesta consiste en una página web que utiliza la distancia Damerau Levenshtein y lemmatizer para llevar las palabras escritas y con errores ortográficos a su forma raíz y aplican una técnica conocida como NER (Reconocimiento de Entidades Nombradas por sus siglas en inglés) que se basa en de identificar palabras claves mediante el entrenamiento de un modelo con 1500 ejemplos de oraciones con entidades etiquetadas para hallar indicios de narcotráfico. Luego, el Sistema de gestión y búsqueda de sentencias judiciales (Gummy[13]) entrena un NER para poder identificar nuevas entidades en el ámbito legal como personas y organizaciones para detectar aquellas entidades de índole legal más relevantes. Por otro lado, Madrazo[14] propone un reconocimiento de entidades en relatos de denuncias, en donde se entrenó un modelo NER para poder identificar

entidades como pueden ser personas, patentes, dni, teléfono y barrio para poder agilizar los trámites relacionados a las denuncias. Nuestro trabajo se encarga de la clasificación de fallos de la Corte Suprema de Justicia de la Nación. Para ello se hace uso de técnicas como reglas heurísticas y un modelo de Machine Learning entrenado para extraer aquellas palabras que dan un indicio de a qué categoría puede pertenecer un fallo determinado.

### 3 Preparación del Dataset de Entrenamiento

Las técnicas de IA como las utilizadas en este trabajo, se basan en el análisis de grandes volúmenes de información que se utilizan para desarrollar modelos que “aprendan” y luego puedan predecir información faltante basados en dicho aprendizaje. Para esto la preparación de la información, o *dataset* es clave para poder obtener los mejores resultados posibles en el aprendizaje de un modelo. En este sentido, es muy importante utilizar datos lo más representativos posibles y que le permitan al modelo determinar las características más relevantes para poder realizar la clasificación. Es por esto que se utilizaron fallos de un período completo (2021) para poder capturar los diferentes elementos representativos.

Los documentos empleados en este dataset son los fallos de la Corte Suprema de Justicia de la Nación. Este dataset se completa con una planilla de cálculos que caracteriza los fallos, y contiene información vital extraída de los mismos. Esta planilla se utilizó para el entrenamiento y validación de modelos. De las 67 columnas que contiene, resultan de interés para el entrenamiento las siguientes: (i) Causa, (ii) materia, (iii) Vía de acceso. La planilla fue realizada por investigadores del Instituto de Derecho Procesal mediante un formulario de Google. Parte de la misma puede verse en la Figura 1. Se contemplaron 8192 fallos de los cuales finalmente se utilizaron 5097. Este filtrado se realizó dado que no se podía acceder al documento PDF de algunos fallos. Por otra parte, en el dataset las causas son muy desparejas respecto de las materias; por ejemplo, algunas materias tienen solo 500 fallos, mientras que otras materias tienen varios miles. Por este motivo, para obtener un dataset equilibrado para entrenar al modelo se tomaron 500 fallos por materia (el mínimo número de fallos para una categoría).

Como paso previo al entrenamiento, se aplicaron distintos preprocesos para aumentar su calidad. Las técnicas aplicadas fueron el pasaje del texto a minúscula, eliminación de *stopwords*, y lematización o *stemming*, y se detallan a continuación

Causa	Fecha resolución anali	Cantidad expedientes	Cantidad de causas	Materia	Vía de acceso
CAF 017171/2014/CSJ	8/4/2021	1	1	Contencioso Administri	Competencia apelada (recursos)
:CNT 005450	23/3/2021	1	1	Laboral	Competencia apelada (recursos)
:CNT 062006/2016/1/R	23/3/2021	1	1	Laboral	Competencia apelada (recursos)
:CNT 012945/2012/1/R	23/3/2021	1	1	Laboral	Competencia apelada (recursos)
:CNT 025017/2018/1/R	23/3/2021	1	1	Laboral	Competencia apelada (recursos)
:CNT 004207/2016/1/R	23/3/2021	1	1	Laboral	Competencia apelada (recursos)
:CNT 027304/2013/2/R	23/3/2021	1	1	Laboral	Comoetencia aovelada (recursos)

Fig 1. Una parte de la planilla utilizada como dataset.

Primero, se reescribieron los fallos completamente en minúscula para simplificar la escritura de reglas que comparan ciertos bloques de texto. Luego, como existen una amplia variedad de palabras que carecen de significado por sí mismas (conocidas como *stopwords* en inglés) se realizó un proceso de filtrado para eliminarlas. Dado que estas palabras son muy frecuentes, mantenerlas puede producir que el algoritmo se confunda y falle en la predicción. Por ejemplo, para las siguientes frases: “Tribunal en lo contencioso administrativo” y “Tribunal en lo civil”, el algoritmo podría observar que ambas frases tienen las palabras en común “Tribunal”, “en”, “lo” y esto podría hacer que el modelo clasifique a ambas etiquetas en una misma categoría ya que tienen tres palabras en común entre los dos ejemplos y se parecen (por ejemplo, podría clasificar un fallo como contencioso administrativo cuando es civil). Por otro lado, caracteres especiales, como puede ser un salto de línea, podría llegar a generar ruido al modelo en el momento de la clasificación y en el peor caso podría tomarlo como un parámetro más al momento de clasificar. Por último, dado a que existen un conjunto de palabras que pueden aparecer a lo largo de un texto de forma conjugada, se necesita una técnica para encontrar una forma ‘estándar’ para obtener tener una sola representación de dicha palabra y por ende facilitar aún más la identificación de las características de cada fallo. Para poder resolver esto último se pueden utilizar técnicas como stemming o lemmatization que buscan como representación, una palabra raíz. Por ejemplo, si se tiene la siguiente oración: “Las flores florecían hermosamente en el jardín”, al aplicar stemming el resultado es “Las flor flore herm en el jard” mientras que si aplicamos lemmatization el resultado es: “La flor florecer hermosamente en el jardín”. Si bien ambos buscan las raíces de las palabras, se puede observar que “flore”, “jard” y “herm” no son palabras válidas mientras que en el caso de lematización todas las palabras sí lo son. De estas dos, se prefirió utilizar lemmatization, dado que permite obtener una palabra, mientras que stemming trabaja con partes de palabras que podrían causar ruido e inconsistencias que comprometen al algoritmo.

#### 4 Detalle de la Propuesta

El algoritmo propuesto está escrito en el lenguaje de programación Python y utiliza diferentes librerías como SpaCy [5], NLTK[6] para PLN, mientras que para entrenamiento de modelos de IA se utilizó la librería Keras[7]. El procedimiento de dicho algoritmo se basa en la aplicación de varias reglas heurísticas con PLN y un modelo de aprendizaje supervisado para complementar aquellos casos en los que las reglas no pueden cubrir. El algoritmo itera por todos los fallos, y los clasifica por **Materia** (Civil y Comercial, Contencioso Administrativo, Electoral, Previsional, Penal, Laboral, Familia, Otro), **Origen** (Provincial o Federal) y **Vía de Acceso** (Competencia apelada de recursos, Juicio originario, por Competencia). Como resultado final, el algoritmo muestra en números cuántos fallos pertenecen a cada categoría mencionada anteriormente, la cantidad de aciertos y el porcentaje de precisión para la clasificación por Materia y por Vía de Acceso.

El algoritmo final implementado se puede resumir en tres etapas: (i) clasificación por materia, (ii) clasificación en base al ámbito federal / provincial y (iii) clasificación por vía de acceso. La Figura 2 resume el procedimiento.

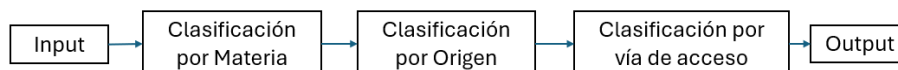


Fig 2. Resumen del procedimiento

El resto de la subsección se divide de la siguiente manera. La subsección 4.1 describe el entrenamiento y selección del modelo de Machine Learning empleado para complementar las reglas heurísticas del algoritmo planteado. La subsección 4.2 describe la primera parte del proceso del algoritmo de clasificador donde se clasifica por materia. La subsección 4.3 continúa la descripción del algoritmo de clasificación por origen. Por último, la subsección 4.4 describe la clasificación por vía de acceso.

#### 4.1 Modelo de Machine Learning para clasificar por Materia

Dado a que no toda la información se puede extraer por reglas, fue necesario entrenar un modelo de aprendizaje supervisado para los casos en los cuales las reglas no son capaces de obtener alguna pista sobre la clasificación a la materia a la que pertenece un fallo dado. Para ellos se utilizó el dataset explicado en la sección 3 y se realizó un script específico con Keras para su creación, entrenamiento y posterior almacenamiento. Se dividió el dataset de la siguiente forma: un 80% como datos de entrenamiento y un 20% como datos de *testing*.

A partir de los datos suministrados por la planilla dada (ver Figura 1), se procedió a leer cada fallo almacenado como PDF con su etiqueta determinada anteriormente por un experto en la materia. A cada fallo se le aplicó el preprocesamiento en 3 etapas descripto en la sección 3: transformación del texto completo a minúscula, eliminación de stopwords y caracteres especiales como el salto de línea, y lematización para poder tener una única representación de palabras que pueden llegar a tener múltiples conjugaciones.

Para elegir el modelo adecuado, se realizaron pruebas con diferentes modelos, y entre ellos el modelo LSTM superó ampliamente al resto. Se utilizaron las métricas estándar para medir el rendimiento: accuracy, recall, y F-score. El cálculo de la métrica de accuracy representa la cantidad de ejemplos correctamente clasificados sobre el total de los ejemplos, mientras que el recall nos indica cuántos ejemplos que son clasificados correctamente de una categoría son clasificados correctamente. Por último F-score es otra métrica que combina accuracy y recall ponderando las de una manera más balanceada para tener una visión más general. Ver Tabla 1.

**Tabla 1.** Algunas pruebas realizadas con algunos modelos y sus resultados.

Modelo	Parametros	Resultado
RNN	<i>Learning Rate = 0,01. Epoch = 50.</i> <i>F. Activación= Softmax</i>	Accuracy:0.71
		F-score:0.72
		Recall:0.74
		Accuracy:0.55
Random Forest	<i>N = 100</i>	F-score:0.54
		Recall:0.56
		Accuracy:0.62
Árboles de decisión		F-score:0.61
		Recall:0.61
		Accuracy:0.25
KNN	<i>N = 6</i>	F-score:0.17
		Recall:0.23
		Accuracy:0.83
LSTM	<i>Learning Rate = 0,01. Epoch = 50.</i> <i>F. Activación= Softmax</i>	F-score:0.82
		Recall:0.82

El modelo utilizado es un modelo secuencial que tiene las siguientes capas: (i) Embedding: Capa de transformación de cadenas de texto en números. (ii) LSTM: Modelo de redes recurrentes empleado para la clasificación. (iii) DropOut. (iv) Batch Normalization. (v) Flatten y (vi) Dense.

#### 4.2 Reglas heurísticas para clasificar por Materia

Para realizar la clasificación, el algoritmo hace pasar cada fallo por una serie de reglas que extraen características que ubican a los fallos en diferentes las diferentes materias. Las materias son: **Civil y Comercial, Contencioso Administrativo, Electoral, Previsional, Penal, Laboral, Familia**. En el caso en que algún fallo pase por todas las reglas y no haya sido clasificado, se lo pasa al modelo de inteligencia artificial para que este entregue una predicción y así poder clasificarlo. Por último, en el caso en que el fallo por alguna razón no se haya podido identificar dentro de una materia determinada existe una última categoría llamada **Otros**. Los escritos que quedan en esta categoría deben ser clasificados por un operador humano manualmente.

Las reglas son las siguientes:

- A) Por Nomenclado: existen algunas nomenclaturas que ya ubican el fallo en una categoría dada a partir de los tres primeros caracteres en mayúscula (Ver Tabla 2)

**Tabla 2.** Nomenclatura con categorías.

Primeros Caracteres	Categoría
CIV, CCF, COM	Civil y Comercial
CAF	Contencioso administrativo / tributario
CNE	Electoral
CSS	Previsional
CPE, CFP, CCC, CPF, CPN	Penal y contravencional
CNT	Laboral

Como no todos los fallos respetan esta nomenclatura, son necesarias otras reglas adicionales. Por ejemplo, los escritos que tienen origen en alguna provincia y llegan a la Suprema Corte de Provincial tienen una nomenclatura que no indica ninguna materia como pueden ser FCB (Justicia Federal de Córdoba), FLP (Justicia Federal de La Plata), entre otros.

- B) Para el caso en que los escritos sean originarios de la Corte Suprema de Nación: Se busca en la carátula algún fallo cuya nomenclatura se corresponda con la regla A.
- C) En el caso de que el escrito sea originario de CABA: Se extrae información del tribunal. Esta información extraída nos indica la materia que maneja dicho tribunal y por ende se puede identificar la materia del fallo.
- D) En el caso de los fallos que son originarios de las provincias: (i) se extrae en la carátula algún otro expediente que cumpla con la regla A. (ii) En el caso que no se encuentre un expediente que cumpla con la regla A, se fija en los órganos intervinientes. Por ejemplo, si en el fallo participa ANSES la materia será previsional, en cambio si participa AFIP la materia será contencioso administrativo / tributario.

Si el fallo no cumple con las reglas anteriores, es procesado por el algoritmo de inteligencia artificial para determinar la materia del mismo. En el caso que el fallo no pueda clasificarse en las materias anteriores, el modelo de machine learning lo va a marcar como 'Otros' para que en un futuro dicho fallo sea clasificado por un experto.

### **4.3 Reglas heurísticas para clasificar según el fallo sea Provincial o Federal**

Luego de finalizada la anterior clasificación, el algoritmo entra en la etapa de clasificación de fallos según si el fallo se origina en provincia o es federal. Este proceso se compone únicamente de reglas de extracción para poder clasificar los fallos.

- A) Por nomencorado: Los fallos tienen en la nomenclatura aclarando si es federal o no.
- B) Si la causa es contra organismos del estado nacional, el fallo es federal.
- C) Si el fallo pertenece a la materia previsional y contencioso administrativo/tributario, el fallo es federal
- D) Si el fallo va en contra de una ley federal, contra la constitución o contra los tratados internacionales, el escrito es federal
- E) El escrito es provincial en general si el último tribunal interviniente es el Superior Tribunal de la provincia

El resto de los fallos serán considerados como fallos de provincia.

#### 4.4 Reglas heurísticas para clasificar según vía de acceso.

Por último, el algoritmo clasifica los fallos por vía de acceso de competencia apelada de recursos, por competencia y por juicios originarios. Nuevamente, se basan en reglas heurísticas para determinar a qué categoría pertenece.

- A) Si aparece la leyenda 'ORIGINARIO' en la carátula del fallo, entonces el fallo pertenece a la categoría de Juicios Originarios
- B) Si aparece la leyenda 'Competencia', en la carátula del fallo, entonces el fallo pertenece a la categoría de Competencia.
- C) Si en la parte final de la nomenclatura aparece 'CS' o 'RH', entonces el fallo pertenece a la categoría de Competencia apelada de Recursos
- D) Si en el cuerpo del fallo se hace referencia a Recurso de Queja o Queja REF, el fallo pertenece a la categoría de Competencia apelada de Recursos.

## 5 Resultados

Se programó un script en Python aplicando las reglas y complementando con el modelo de inteligencia artificial entrenado. Luego se pasaron por 5097 archivos PDF de distintos fallos de los cuales se filtraron los escritos que son dictámenes. Como resultado se obtuvieron las tasas de acierto que se muestran en la Figura 3.

```
Civil y Comercial: 494, Contencioso Administrativo: 1173, Electoral: 10, Previsional: 610,
Penal: 2263, Laboral: 574, "Familia": 2 Sin ID: 0
Aciertos: 4705, Accuracy: 0.9178696839641045
Casos federales: 2034, Casos Provinciales: 2025
Aciertos: 3818, Accuracy: 0.9406257698940625
Clasificación por vía: Por Competencia: 834, Por Origen: 180, Por Recurso: 4059, Otro: 53
Aciertos: 5014, Accuracy: 0.9781506047600468
```

Fig 3. Imagen del output de la herramienta.

Los resultados obtenidos son prometedores. Se consiguió un 91% de accuracy del algoritmo en cuanto a la clasificación de fallos por materias. Además se consiguió un

97% de accuracy en cuanto a clasificación en base a la vía de acceso. Y finalmente, con respecto a la clasificación por origen se obtuvo un 94% de accuracy.

Debido al hecho de que existen algunos fallos que necesitan la interacción de un historial de fallos para poder saber la materia, es necesario seguir refinando la herramienta para que los pueda cruzar y así mejorar aún más el porcentaje de precisión. Al mismo tiempo, es necesario seguir mejorando reglas para depender menos de un modelo de inteligencia artificial, ya que si bien tiene un buen porcentaje de precisión siempre está la posibilidad de que se equivoque. Como los resultados van a ser analizados por expertos del dominio para sacar estadísticas, se necesita tener el mayor porcentaje de aciertos para que estos puedan obtener estadísticas más precisas.

Con respecto a los recursos, se obtuvieron muy buenos resultados. Se analizaron aquellos errores en los que no predecía de forma correcta, y en estos se encontró que si bien hay algunos en los que realmente se confunde, existieron otros en los cuales no cumplían con ninguna regla, lo cual se puede seguir mejorando aún más dichas reglas.

Por último, con respecto a la clasificación de los fallos según su origen, se puede continuar haciendo más pruebas y descubriendo nuevas reglas para mejorar todavía más el índice de acierto de la misma.

## 6 Conclusiones

Las herramientas de IA y PLN tienen un gran potencial en la producción de indicadores de funcionamiento y rendimiento del servicio de justicia. Estas herramientas automatizan la lectura de los documentos para extraer cierta información lo cual permite dotar de agilidad y eficiencia a la elaboración de dichos indicadores. Si bien algunas de las herramientas son probabilísticas, los resultados obtenidos hasta el momento, superan en todos los casos el 90% de *accuracy*, lo cual es un valor muy alto, considerando que el trabajo humano también es propenso a errores. Es así que los resultados son muy prometedores, y de todas formas, el equipo ya ha identificado diferentes puntos de mejora para mejorar aún más los niveles de *accuracy*.

Con esta asistencia, es posible enfrentar uno de los inconvenientes que tradicionalmente han obstaculizado el desarrollo de una práctica consistente de medición de resultados en el servicio de justicia. Suele considerarse que la producción de información pública de calidad sobre la gestión del sistema de justicia, requiere asignar recursos del poder judicial (o de las oficinas gubernamentales que participan en el diseño de políticas públicas judiciales), a tareas ajenas a su núcleo de incumbencia. En ausencia de una cultura arraigada de rendición de cuentas, existen estímulos constantes para aplicar esos esfuerzos humanos y presupuestarios en finalidades distintas a la producción de información que finalmente servirá para supervisar el funcionamiento de quien la pone a disposición. La eficiencia que exhibe el uso de herramientas de IA y PLN permite mitigar esta clase de reparos basados en la escasez presupuestaria y en la asignación de prioridades en la agenda de trabajo de los agentes del poder judicial u oficinas competentes en la planificación y gestión del servicio de justicia.

Mejorar la transparencia y el acceso a la información sobre la actuación de los tribunales es condición esencial para dotar de mayor legitimidad al sistema de justicia y enriquecer la calidad de la discusión pública tendiente a reformar sus prácticas y

procedimientos. Sin datos de calidad, ninguna de dichas aspiraciones es realista, por lo que corresponde avanzar en el diseño e implementación de esta clase de herramientas tan relevantes para fortalecer la calidad institucional y la tutela efectiva de los derechos.

## 7 Trabajo Futuro

Es importante continuar mejorando la precisión y grado de acierto de los algoritmos aplicados y avanzar progresivamente hacia la obtención de información de calidad sobre otros aspectos de la actuación jurisdiccional de la Corte Suprema. Los resultados obtenidos hasta el momento ya muestran potencial para convertirse en herramientas de utilidad real en el objetivo final de producir estadísticas de calidad. No obstante, los porcentajes de precisión en sus diferentes métricas pueden mejorarse mediante el ajuste del algoritmo existente o con la ayuda de técnicas complementarias.

Otra mejora que se plantea es trabajar en la forma de detectar personas físicas y jurídicas mediante el entrenamiento de un NER (Named Entity Recognition). Ya que identificar las personas en pleito pueden brindar pistas e indicios para poder identificar a qué categoría pertenece un determinado fallo y por ende mejorar la clasificación de materia. Por un lado pareciera simple identificar personas físicas y jurídicas, dado que existen organismos que consolidan la información: Registro de las Personas y Registro de las sociedades. Sin embargo, ya sea por errores, por falta de sincronismo o por simplemente estilos de redacción, una búsqueda precisa no siempre es posible. Es por ello que entrenar un NER, el cual reconoce ciertas entidades a partir de su contenido y el contexto es una herramienta muy útil.

Finalmente, se pretende obtener información adicional, que en relación con la información ya obtenida mejoraría los resultados generales. Específicamente, se plantea clasificar los fallos que son recursos en base a si son REF o Queja REF, al igual que poder realizar una clasificación de los recursos en base a la resolución que toma la CS en base a si son definitivas, interlocutorias o simples.

## References

1. V. Giannini, L. (2018). La producción jurisdiccional de la Corte Suprema en su actual integración. En: Revista Anales de la Facultad de Ciencias Jurídicas y Sociales de la Universidad Nacional de La Plata (48), pp. 1208-1259. Disponible en: <https://revistas.unlp.edu.ar/RevistaAnalesJursoc/article/view/5293>;
2. Giannini, L. (2020). La producción jurisdiccional de la Corte Suprema de Justicia de la Nación. En: Revista Anales de la Facultad de Ciencias Jurídicas y Sociales de la Universidad Nacional de La Plata (50), pp. 791-832. Disponible en: <https://doi.org/10.24215/25916386e067>;
3. Giannini, L. (2021). La Corte Suprema. Actualidad. Funcionamiento. Propuestas de reforma. Buenos Aires: Editores del Sur;
4. Giannini, L. (2022). La producción jurisdiccional de la Corte Suprema de Justicia de la Nación (2020-2021). Revista de Derecho Procesal, Santa Fe: Rubinzal Culzoni, n° 2022-2, pp. 489-578.

5. Spacy, “Spacy: Industrial-Strength Natural Language Processing,” 2023. <https://spacy.io/>, last accessed 2024/03/25
6. NLTK, Natural Language Tool Kit, <https://www.nltk.org/>, last accessed 2024/03/25.
7. Keras, <https://keras.io/>, last accessed 2024/03/25.
8. Kit de herramientas global sobre IA y el estado de derecho para el poder judicial, UNESCO, nro de catalogo 0000387331,2023
9. C. Serventich, «Inteligencia artificial en el proceso penal: ¿Más vale humano conocido o algoritmo por conocer?», RJA, vol. 3, n.º 2, pp. 869-880, dic. 2022.
10. Bustos, Lex & Aguerre, Carolina. (2021). Preparación del sector judicial para la inteligencia artificial en América Latina (compilación).
11. Sposito, O. M., Ryckeboer, H. E., Ledesma, V., Procopio, G., Matteo, L., Gargano, C., ... & Pérez Villar, G. (2021). Propuesta para la construcción de un Corpus Jurídico utilizando Expresiones Regulares. In XXVII Congreso Argentino de Ciencias de la Computación (CACIC)(Modalidad virtual, 4 al 8 de octubre de 2021).
12. NLP aplicado a análisis de texto, Alias Gerardo Cassanelli, Rodrigo dirigido por Ana Haydee Di Iorio, Universidad Nacional de Mar del Plata, Facultad de Ingeniería (2019)
13. Sistema de gestión y búsqueda de sentencias judiciales utilizando procesamiento del lenguaje natural, Gumy, Juan Nicolás dirigido por Ing. Ariel Podesta y codirigido por Ing. Bruno Constanzo, Universidad Nacional de Mar del Plata, Facultad de Ingeniería (2021).
14. Reconocimiento de Entidades en Relatos de Denuncias mediante Aprendizaje Automático y Procesamiento de Lenguaje Natural, Gustavo Ariel Rivera Madrazo, dirigido por Dr. Eduardo Xamena, Universidad Católica de Salta, Facultad de Ingeniería.