

Septiembre 2024

Datos



Usos inteligentes de datos en el Estado

Juan Pablo Ruiz Nicolini
Mariana Kunst
Juan Manuel Dias



fundar

Usos inteligentes de datos en el Estado

Juan Pablo Ruiz Nicolini

Mariana Kunst

Juan Manuel Dias

- Transformar el Estado
- Generar riqueza
- Promover el bienestar



Sobre Fundar

Fundar es un centro de estudios y diseño de políticas públicas que promueve una agenda de desarrollo sustentable e inclusivo para la Argentina. Para enriquecer el debate público es necesario tener un debate interno: por ello lo promovemos en el proceso de elaboración de cualquiera de nuestros documentos. Confiamos en que cada trabajo que publicamos expresa algo de lo que deseamos proyectar y construir para nuestro país. Fundar no es un logo: es una firma.

En Fundar nos dedicamos al estudio e investigación de políticas públicas sobre la base de evidencia. Como parte de nuestra política de promover la transparencia y promoción de la discusión pública, disponibilizamos los datos utilizados para nuestros análisis, para que cualquier persona que lo desee pueda replicar los análisis realizados y generar nuevas investigaciones

En Fundar creemos que el lenguaje es un territorio de disputa política y cultural. Por ello, sugerimos que se tengan en cuenta algunos recursos para evitar sesgos excluyentes en el discurso. No imponemos ningún uso en particular ni establecemos ninguna actitud normativa. Entendemos que el lenguaje inclusivo es una forma de ampliar el repertorio lingüístico, es decir una herramienta para que cada persona encuentre la forma más adecuada de expresar sus ideas.

Trabajamos en tres misiones estratégicas para alcanzar el desarrollo inclusivo y sustentable de la Argentina:

Generar riqueza. La Argentina tiene el potencial de crecer y de elegir cómo hacerlo. Sin crecimiento, no hay horizonte de desarrollo, ni protección social sustentable, ni transformación del Estado. Por eso, nuestra misión es hacer aportes que definan cuál es la mejor manera de crecer para que la Argentina del siglo XXI pueda responder a esos desafíos.

Promover el bienestar. El Estado de Bienestar argentino ha sido un modelo de protección e inclusión social. Nuestra misión es preservar y actualizar ese legado, a través del diseño de políticas públicas inclusivas que sean sustentables. Proteger e incluir a futuro es la mejor manera de reivindicar el espíritu de movilidad social que define a nuestra sociedad.

Transformar el Estado. La mejora de las capacidades estatales es imprescindible para las transformaciones que la Argentina necesita en el camino al desarrollo. Nuestra misión es afrontar la tarea en algunos aspectos fundamentales: el gobierno de datos, el diseño de una nueva gobernanza estatal y la articulación de un derecho administrativo para el siglo XXI.

Cita sugerida

Ruiz Nicolini, J. P., Kunst, M., y Dias, J. M. (2024). [Usos inteligentes de datos en el Estado](#). Fundar.

Licencia

Esta obra se encuentra sujeta a una licencia [Creative Commons 4.0 Atribución-NoComercial-Sin-Derivadas Licencia Pública Internacional \(CC-BY-NC-ND 4.0\)](#). Queremos que nuestros trabajos lleguen a la mayor cantidad de personas en cualquier medio o formato, por eso celebramos su uso y difusión sin fines comerciales.

Agradecimientos

Agradecemos especialmente a Paula Luvini, Daniela Risaro y Daniel Yankelevich por sus aportes para la elaboración de este documento.

Índice

Usos inteligentes de datos en el Estado	5	<u>Introducción</u>
	6	<u>Alerta temprana de interrupción de trayectorias escolares en Mendoza</u>
	10	<u>Sistema Estadístico de Población</u>
	14	<u>Infraestructura pública digital para procesos electorales</u>
	18	<u>Clasificación y apertura de datos de sentencias judiciales vinculadas a casos de violencia de género</u>
	22	<u>Segmentación de la población objetivo en el diseño de políticas culturales</u>
	26	<u>Análisis de comportamiento turístico a partir de <i>big data</i></u>
	30	<u>Conclusión</u>
	31	<u>Bibliografía</u>



Introducción

En la actualidad, la información se ha convertido en un recurso estratégico y el Estado es uno de los principales productores de datos del país. A su vez, los datos son un insumo fundamental para el diseño de políticas públicas más eficientes y eficaces, por lo que es imprescindible fomentar tanto su cuidado como su uso. La comprensión y el manejo adecuado de los datos contribuyen a mejorar su calidad y fortalecer la confianza de la información derivada de ellos.

Partiendo de este diagnóstico, desde Fundar trabajamos un conjunto de propuestas que consideramos no pueden faltar en la discusión y definición de una estrategia de datos para la administración pública ([Luvini et al. 2023](#)). A través de ellas, se busca promover una mejora en la información disponible para la formulación de políticas públicas. Sostenemos que un estado inteligente es aquel que sabe, recuerda, razona, planifica y pone en acción sus planes con datos.

A lo largo de los últimos años, observamos avances significativos en el Estado nacional en términos de digitalización y automatización de trámites y de procesos internos, así como en la implementación de agendas relacionadas con la protección de datos personales, acceso a la información pública y políticas de gobierno abierto.

Este documento presenta algunos de esos casos innovadores de uso de datos diferentes áreas de política pública. Todas las experiencias fueron desarrolladas en años recientes, muchas de ellas en el contexto de la pandemia de COVID-19. Muchas fueron también pioneras en los ámbitos de su implementación. Dan cuenta de una variedad de intervenciones en las áreas de diseño y gestión de políticas públicas en la conservación, el aprovechamiento, la disponibilización y acceso público a los datos.

El objetivo de esta presentación es mostrar el potencial transformador de las políticas de uso de datos.

Con esta intervención buscamos crear conciencia sobre la importancia del trabajo con datos, promover el compromiso con su sostenibilidad en la comunidad de datos del Estado, inspirar a los equipos de diversas dependencias y fomentar una cultura de datos tanto a nivel nacional como subnacional.

→ Alerta temprana de interrupción de trayectorias
escolares en Mendoza



Las contribuciones a la redacción del caso se basaron en entrevistas moderadas por el equipo de datos de Fundar con las personas responsables del diseño e implementación del SAT, como Romina Paola Durán (Dirección de Evaluación de la calidad educativa - Mendoza), Esteban Torre, Carla Papparella, Juan Xanthopoulos (CIPPEC) y Matías Frydman (MINCyT-UBA-CID).



Alerta temprana de interrupción de trayectorias escolares en Mendoza



El Sistema de Alerta Temprana (SAT) es un sistema predictivo para prevenir interrupciones de trayectorias escolares de estudiantes secundarios en la provincia de Mendoza. El proyecto se propone como una herramienta de soporte para la gestión educativa que permite potenciar el acompañamiento a las trayectorias de los y las estudiantes, analizar e identificar las causas que las debilitan e implementar acciones para sostener y fortalecer su escolaridad.

Mendoza fue la primera provincia del país en implementar un sistema de alertas tempranas en educación. En parte, esto fue gracias a las capacidades aportadas por el portal Gestión Educativa Mendoza (GEM), una fuente de datos ordenada proveniente del sistema de información educativo de la provincia.

La implementación de este sistema de alertas tempranas surge de la necesidad de la Dirección General de Escuelas (DGE) de automatizar procesos internos y se inspira en experiencias latinoamericanas como la del [Ministerio de Educación de Guatemala \(2019\)](#).

Esta herramienta de gestión agiliza la identificación de estudiantes en riesgo de deserción escolar y permite la intervención temprana para evitar que se interrumpan sus trayectorias escolares.

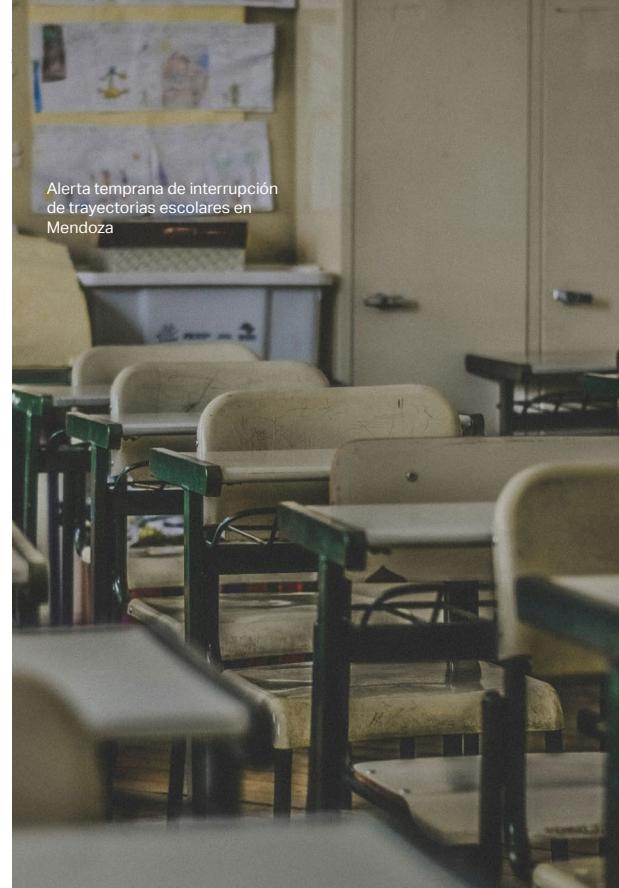
Por qué importa

- El sistema educativo registra mucha información y no se aprovechaba todo su potencial.
- Las escuelas destinaban tiempo y recursos para identificar alumnos en riesgo de abandono con un seguimiento caso por caso, sin ninguna sistematización ni agregación de la información.
- Las escuelas disponen de más tiempo y recursos para diseñar estrategias de intervención que prevengan que los estudiantes interrumpan su trayectoria escolar.

Cómo funciona

El SAT es un proyecto interdisciplinario que involucra a distintas instituciones: la Dirección de Evaluación de la calidad educativa de la DGE de Mendoza, el Centro Interinstitucional de Ciencia de Datos del Ministerio de Ciencia y Técnica y la UBA (MINCyT-UBA-CID) y el Programa de Educación de CIPPEC. El proyecto también contó con la asistencia técnica del Banco de Desarrollo de América Latina (CAF). En la última instancia, las escuelas trabajaron directamente con el equipo de CIPPEC para analizar cómo intervenir frente a los distintos escenarios de riesgo.

El equipo del LIAA fue el encargado de estimar el riesgo de abandono para cada estudiante. Para ello se diseñó un modelo de aprendizaje automático que predice la no matriculación de los alumnos al año siguiente. El modelo, para el cual se utilizó el algoritmo *Catboost*¹, tiene como objetivo predecir la variable “interrupción de trayectorias escolares” que identifica a los alumnos que están matriculados un año y devuelve una probabilidad de no matriculación de cada estudiante al año siguiente. Para visualizar los resultados se incorporó un módulo de alerta temprana que presenta la probabilidad de no matriculación estimada representada mediante un sistema de semáforo de tres categorías: riesgo bajo (color verde), riesgo medio (color amarillo) y riesgo alto (color rojo).



El módulo Sistema de Alerta Temprana (SAT) permite:

- Visualizar los resultados a través de un módulo de alerta temprana que presenta la probabilidad de no matriculación estimada representada mediante un sistema de semáforo de tres categorías: riesgo bajo (color verde), riesgo medio (color amarillo) y riesgo alto (color rojo).
- Acceder a un listado de posibles causas de abandono y a un registro de estrategias para revertirlo.
- Llevar un registro de intervenciones una vez que se identifica qué estudiante está en riesgo (sólo a los directores y supervisores de las escuelas).
- Tener información actualizada tres veces al año sobre inasistencias y calificaciones del alumno para analizar los cambios en los niveles de riesgo de cada estudiante.
- Descargar las guías de usuario para equipos directivos y para supervisores.

¹ Se utilizó el algoritmo CatBoost debido a su capacidad para manejar diferentes tipos de variables, datos faltantes y su buen desempeño en problemas similares ([Porkhorenkova et al. 2018](#)).

De dónde se obtienen los datos

El modelo que predice la no matriculación utiliza datos de factores relacionados con el alumno, su grupo familiar (como el nivel educativo de los padres) y el establecimiento escolar (características de la escuela) que suman 68 variables en total. Estos datos provienen del GEM, de bases de datos de la Dirección de Evaluación de la Calidad Educativa, de las pruebas Aprender y de relevamientos anuales del Ministerio de Educación de Mendoza. La interoperabilidad se realizó mediante mecanismos oficiales y la DGE aseguró la confidencialidad de la información de cada estudiante durante la implementación del proyecto.

Consideraciones finales

- Mendoza contaba con datos bien organizados, lo que facilitó significativamente la implementación del SAT. Esta disponibilidad de datos estructurados proporcionó una base sólida para el desarrollo y puesta en marcha del Sistema de Alerta Temprana (SAT), permitiendo una mayor precisión en la identificación de posibles casos de abandono escolar y la implementación de estrategias de prevención adecuadas.
- Durante el proceso de implementación una parte fundamental del éxito del proyecto fue el hecho de que todas las instituciones involucradas respetaron las necesidades específicas de la provincia.
- En lugar de proponer soluciones genéricas o un enfoque estándar, las instituciones se adaptaron a las particularidades de Mendoza. Esta colaboración estrecha y el respeto por las particularidades locales permitieron que el proyecto se integrara de manera más efectiva tanto a escuelas públicas como privadas.

Seguir leyendo

- Para obtener más información sobre el proyecto consultar [Alertas tempranas para prevenir el abandono escolar: el caso de la Provincia de Mendoza \(Berniell et al., 2023\)](#), una publicación elaborada en el marco de la asistencia técnica de CAF a la Provincia de Mendoza.
- Para obtener más información sobre cómo funciona el SAT y su impacto en la educación, visitar el [sitio web del proyecto](#).
- Para profundizar en la metodología del SAT visitar el documento de CIPPEC Metodología para el desarrollo de un SAT provincial y evaluación de las condiciones de implementación ([Paparella et al., 2023](#)) que describe los pasos necesarios para implementar un proyecto de estas características.
- Para conocer más sobre las características y principales variables asociadas al abandono en el nivel secundario en Mendoza durante el año lectivo 2019 leer [Diagnóstico del abandono escolar en Mendoza \(Delprato et al., 2023\)](#), un estudio cualitativo elaborado por CIPPEC.
- Para conocer el detalle de la designación de responsables de las diversas tareas y del equipo central de la Dirección de Evaluación de la Calidad Educativa que se encargó del proceso de implementación del sistema en las escuelas, se encuentra disponible [Resolución 2023-5649-E-GDEMZA-DGE](#).



Sistema Estadístico de Población



Las contribuciones a la redacción del caso se basaron en intercambios con el equipo responsable del proyecto, incluyendo a Mariano Fagalde, Matías Somoza y Marilen Morales de la Dirección Nacional de Estadísticas de Población - RENAPER.

DATOS EN EL ESTADO
EN POBLACIONES
Uso de registros administrativos



Sistema Estadístico de Población

El Sistema Estadístico de Población representa un cambio significativo en el manejo de los datos poblacionales en Argentina. Proporciona información actualizada sobre aspectos demográficos, socioeconómicos y de desarrollo territorial del país. Esta información se deriva de los registros administrativos del Registro Nacional de las Personas (RENAPER) y abarca temas como distribución y tendencias de nacimientos y fecundidad, migración internacional e interna, estructura poblacional, apellidos y nombres, cambios de género en el DNI y demoras en el acceso al DNI en recién nacidos, entre otros. Estos datos están desagregados hasta el nivel departamental, lo que permite un análisis detallado y una comprensión más completa de la dinámica poblacional del país.

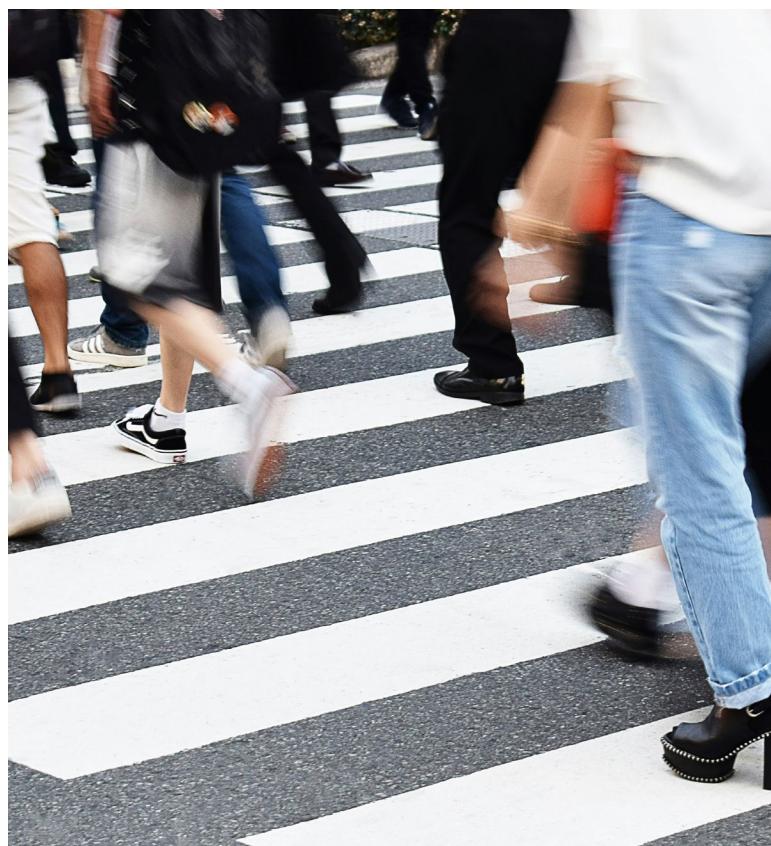
Antes de 2020, los datos del RENAPER rara vez se utilizaban con fines estadísticos y su procesamiento era esporádico, generalmente a cargo del área de sistemas. La base de datos contiene más de 100 millones de trámites históricos, casi 55 millones de personas registradas, más de 52 millones de registros de nacimientos y casi 12 millones de registros de defunciones, entre otros datos. Estas cifras ofrecen una visión general de la base de datos administrativa, que incluye numerosas tablas relacionadas que abarcan diversos momentos históricos del registro y requirieron un esfuerzo considerable de consolidación.

Sin embargo, con la implementación de una base de datos optimizada, derivada de la base de datos administrativa del RENAPER y diseñada específicamente para realizar procesos estadísticos, se agilizó considerablemente el trabajo y se redujeron los errores. Esto permitió que el equipo de la Dirección Nacional de Población (DNP) se enfocara en el análisis y la interpretación de los datos.

En ese mismo año, la DNP introdujo el Sistema Estadístico de Población. A diferencia de otros sistemas basados en encuestas o censos, este recolecta de forma periódica y actualizada los registros administrativos del RENAPER para producir información sociodemográfica de forma periódica y actualizada.

Durante la pandemia el Sistema Estadístico de Población fue utilizado para cruzar información y facilitar la distribución de vacunas.

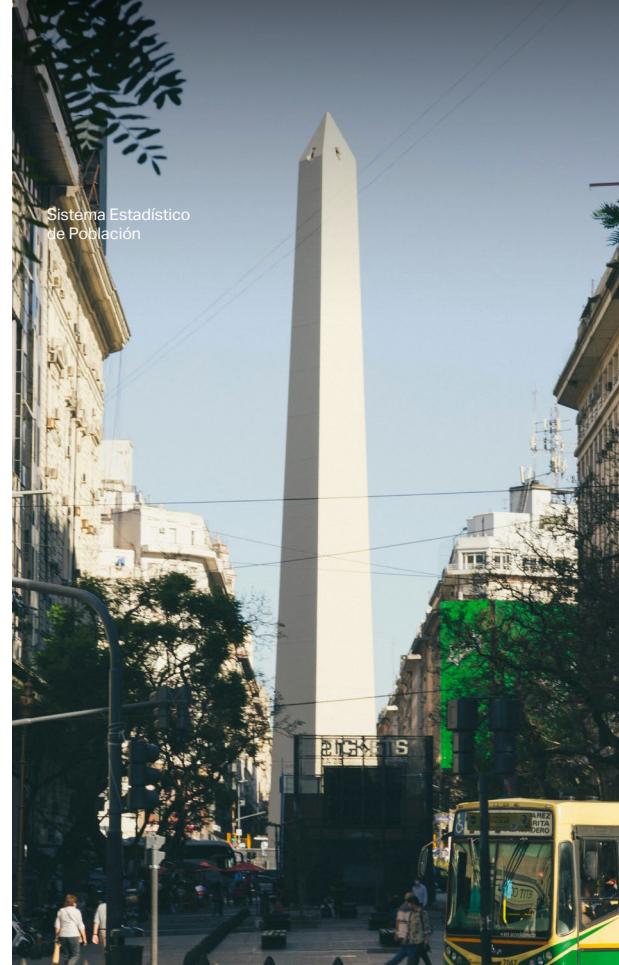
Generó informes específicos sobre la mortalidad según la ubicación geográfica durante la emergencia sanitaria. Además, se analizó el exceso de mortalidad mediante la comparación de las defunciones observadas con las esperadas para ese período del año, basadas en datos históricos. Asimismo, el sistema permitió identificar la distribución espacial de la población de adultos mayores, priorizada en la estrategia de vacunación.



DATOS EN EL ESTADO
EN POBLACIONES
Uso de registros administrativos

Por qué importa

- Existía una gran cantidad y variedad de información proveniente de registros acumulados a lo largo de muchos años de trámites digitalizados. La falta de documentación de dicha información significó empezar prácticamente desde cero.
- Anteriormente, ciertos pedidos de información requerían procesamientos complejos a cargo del área de sistemas que implicaban largos tiempos de respuesta, mayor costo y esfuerzo. En contraste, con el sistema actual las respuestas son más rápidas y se ahorra tiempo y recursos.



Cómo funciona

El sistema opera mediante la recopilación y normalización de datos provenientes del RENAPER. Un equipo de especialistas se encargó de revisar, definir criterios y documentar procesos para transformar un conjunto de datos dispersos en un sistema organizado. Se establecieron métricas específicas para evaluar su calidad, garantizando el cumplimiento de los estándares legales y de protección de datos. El sistema se desarrolló cumpliendo con la [Ley de Secreto Estadístico N°17.622](#) y la [Ley de Protección de Datos Personales N°25.326](#). A su vez, en el año 2022 el RENAPER aprobó su propia política de protección de datos personales.

El Sistema Estadístico de Población permite:

- Acceder de forma abierta a información agregada y anónima a través de diversos medios, como tableros interactivos, estudios, mapas y datos abiertos. Este enfoque garantiza que la información proporcionada no permita la identificación ni la revelación de datos sensibles de las personas.
- Descargar informes elaborados por el Centro Nacional de Estudios de Población.
- Descargar bases de datos del RENAPER en formato CSV.
- Visualizar información espacial y estadística de la población argentina a través de la Infraestructura de Datos Espaciales (IDE).
- Acceder a tableros restringidos dirigidos a audiencias específicas, como Directores de Registros Civiles, que incluyen indicadores de gestión de trámites en su jurisdicción. Estos tableros específicos están protegidos por contraseña y solo son accesibles para usuarios autorizados.

De dónde se obtienen los datos

Los datos provienen de los procesos de solicitud de documentos de identidad como el DNI o el pasaporte. Para garantizar la calidad y la precisión de la información producida, se la contrastó con datos proporcionados por el Ministerio de Salud (en relación con los hechos vitales), registros civiles provinciales y datos del Instituto Nacional de Estadística y Censos (INDEC). Además, se solicitó la colaboración de la Dirección Nacional de Migraciones para realizar cruces de datos y así identificar a las personas que se encuentran fuera del país durante ciertos períodos, lo que permitió filtrar esta información en algunos procesamientos específicos.

Seguir leyendo

- Para obtener más información sobre cómo funciona el Sistema Estadístico de Población, visitar el [portal de datos abiertos](#) que disponibiliza bases de datos para descargar.
- Para acceder a tableros interactivos, estudios, mapas y datos abiertos, visitar el [sitio web](#) de la Dirección Nacional de Población.
- Para visualizar información especial y estadística de la población argentina, visitar el [portal Infraestructura de Datos Espaciales](#).

Consideraciones finales

- La información demográfica que provee este sistema resulta de utilidad para la formulación, monitoreo y evaluación de políticas públicas a nivel nacional y subnacional. Permite obtener una comprensión más clara de la dinámica demográfica en Argentina.
- Fortalece la institucionalidad pública encargada de los asuntos de población y desarrollo a nivel nacional y subnacional, uno de los objetivos del Consenso de Montevideo sobre Población y Desarrollo, Primera Reunión de la Conferencia Regional sobre Población y Desarrollo de América Latina y el Caribe ([N.U. y CEPAL, 2013](#)).
- Aporta transparencia al permitir el acceso a la información pública de la ciudadanía garantizando la confidencialidad de la información personal.
- Se pueden utilizar para mejorar la precisión de los datos de domicilios de las personas, lo cual puede contribuir a mejorar la focalización de las políticas públicas y promover una mayor transparencia en la gestión de los datos administrativos.





Infraestructura pública digital para procesos electorales



Las contribuciones a la redacción del caso se basaron en entrevistas con el equipo responsable del proyecto, incluyendo a Rosario Alessandro, María José Odriozola, Bianca Giacomino, Marcos Buero, Ramiro San Juan y Federico Lazcano, ex miembros del equipo de la Dirección Nacional Electoral - Ministerio del Interior.

Infraestructura pública digital para procesos electorales

El Sistema de Publicación de Resultados Electorales permite el archivo, tratamiento, intercambio y publicación ágil de datos históricos, incluyendo los resultados provisionales de los últimos 10 años de elecciones nacionales, provinciales y locales realizadas bajo la adhesión a ley de simultaneidad de procesos electorales. Permite un análisis detallado y una comprensión más completa de la dinámica poblacional del país, para velar por la consistencia y armonización de los datos, su publicación y resguardo. El proyecto transformó un proceso desordenado en una iniciativa organizada, transparente y accesible.

Antes de 2021, el proceso de recuentos provisionales de votos en Argentina carecía de estándares, normativas o requisitos para la entrega de datos por parte de los proveedores del servicio. Ese año, la Dirección Nacional Electoral (DINE) y el Observatorio Político Electoral del Ministerio del Interior de la Nación definieron y desarrollaron un estándar para subsanar esta inefficiencia y reconstruyeron datos históricos disponibles basados en estas definiciones. El trabajo avanzó en una primera etapa sobre los datos bajo su dominio del período comprendido entre 2011 y 2023, con el objetivo de avanzar en la apertura de datos. Durante este proceso quedó en evidencia la complejidad que implicaba garantizar la integridad de la información, especialmente en lo que respecta a la completitud de los datos provenientes de los recuentos provisarios provistos por las empresas adjudicatarias del operativo electoral. La falta de un método definido para compartir información generó desafíos en relación con la calidad de los datos. El trabajo de análisis de los equipos técnicos de la DINE derivó en el desarrollo de un estándar para velar por la consistencia y armonización de los datos, su publicación y resguardo.

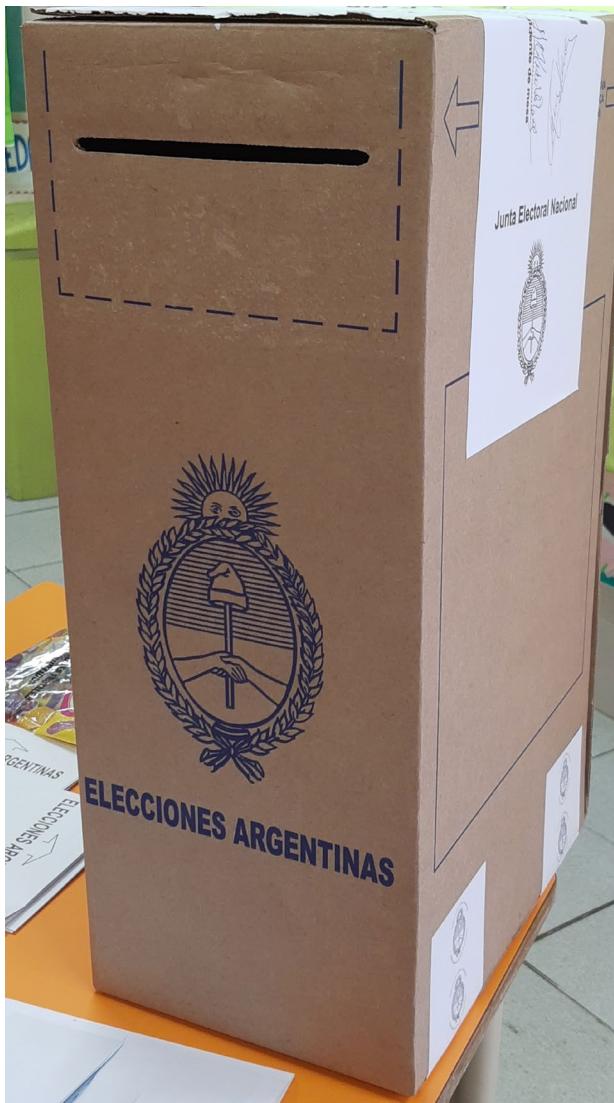


En el contexto de las elecciones presidenciales de 2023, desempeñó un papel relevante para desmentir y combatir las campañas de desinformación.

Al ofrecer datos electorales públicos consistentes, aportó a la transparencia y la integridad del proceso electoral y, de este modo, a preservar la confianza de la ciudadanía en el sistema democrático ([Ávila y Ruiz Nicolini, 2024](#)).

Por qué importa

- Hasta el momento de esta iniciativa, no había antecedentes de la exigencia de un estándar sobre cómo debían entregarse los resultados.
- Disponibiliza los datos de resultados provisionales de los últimos 10 años de elecciones nacionales, modernizando la perspectiva histórica de la democracia.
- Proporciona información esencial para el seguimiento de los resultados electorales y para prevenir posibles amenazas de fraude que pueden socavar la confianza del proceso electoral.



Cómo funciona

El Sistema de Publicación de Resultados Electorales establece un formato de almacenamiento definitivo para los datos electorales, garantizando la preservación de la información con un estándar fácilmente accesible y procesable. Asimismo, facilita la entrega de los datos finales del recuento provisional de resultados de las elecciones nacionales y es adaptable para su uso en diferentes gobiernos y contextos. Por último, facilita el acceso a datos de manera oportuna para diversos públicos.

El sistema permite:

- Acceder a los datos de los resultados de manera desagregada y sistematizada desde las elecciones 2011 hasta el último proceso electoral en 2023. Además facilita la consulta de los resultados provisорios por cada agrupación política, organizados por categorías electorales y ámbitos geográficos, a través de un [buscador avanzado](#) y ofrece una herramienta para generar informes personalizados y la descarga de datos de las consultas generadas.
- Un acceso a través de una [API](#) (o interfaz de programación de aplicaciones).
- Conocer el [Estándar para el almacenamiento y preservación de resultados electorales](#), que detalla los campos y convenciones de uso.
- Obtener la totalidad de los datos en formatos abiertos y reutilizables desde el [portal del gobierno nacional](#).

DATOS EN EL ESTADO EN
PROCESOS ELECTORALES
Estándares de Datos

De dónde se obtienen los datos

El escrutinio provisorio se realiza el día de la elección con los resultados y son informados por cada presidente de mesa al Ministerio del Interior mediante un telegrama. Este telegrama, elaborado y firmado por el presidente de mesa y los fiscales de las agrupaciones, detalla los votos contados. El personal del correo lo transmite a la Justicia Electoral y a la Dirección Nacional Electoral para totalizar y difundir los resultados preliminares². Una parte nodal de todo este proceso reside en que, para cada ciclo electoral, la DINE realiza una licitación para la contratación de un servicio de escrutinio provisorio.

Consideraciones finales

- El proyecto promueve la transparencia y legitimidad del proceso democrático, fortaleciendo así la confianza en el sistema electoral.
- Resulta destacable la formalización normativa de este avance a través de la publicación de la Disposición [DI-2024-1-APN-DNE#MI](#) en el Boletín Oficial de la República Argentina.
- Aunque este proyecto no se pueda trasladar automáticamente a otros ámbitos electorales, como elecciones para cargos provinciales o locales, tiene el potencial de ser adaptado para su aplicación en estos ámbitos.

² Estos resultados no cubren todas las mesas debido a errores o telegramas no enviados, y no tienen validez legal ni sirven para proclamar a los electos. La DINE es la responsable de los datos derivados del recuento provisorio de votos, mientras que los datos del recuento definitivo están bajo la jurisdicción de la Justicia Electoral, que incluye los Juzgados Federales y la Cámara Nacional Electoral.



Seguir leyendo

- Para conocer la herramienta de búsqueda avanzada y generación de informes personalizados basada en los resultados provisionales de elecciones nacionales, provinciales y locales de los últimos años, visitar el [sitio de Resultados Electorales de la DINE](#). Además, se proporcionan especificaciones y acceso para que desarrolladores de software puedan utilizar esta información.
- Para acceder a los conjunto de datos y todos los recursos de la DINE de manera libre y abierta, visitar el [Portal Nacional de Datos Abiertos](#).

→ Clasificación y apertura de datos de sentencias
judiciales vinculadas a casos de violencia de
género



Las contribuciones a la redacción del caso se basaron en entrevistas con el equipo responsable del proyecto, incluyendo a Pablo Casas, Yasmín Belén Quiroga del Juzgado Penal, Contravencional y de Faltas nº 10 de la Ciudad Autónoma de Buenos Aires.



Clasificación y apertura de datos de sentencias judiciales vinculadas a casos de violencia de género



El Juzgado Penal, Contravencional y de Faltas N° 10 de la Ciudad Autónoma de Buenos Aires aborda la problemática de la violencia de género valiéndose de distintas acciones. Una de ellas es la digitalización de expedientes y la implementación de herramientas de inteligencia artificial para la gestión y análisis de datos judiciales. El proyecto AymurAI surge con el objetivo de automatizar la clasificación y análisis de datos de sentencias. Este proyecto permite resolver la falta de datos unificados sobre violencia de género al proporcionar una plataforma que integra y organiza la información de manera sistemática.

La clasificación de datos de sentencias judiciales es novedosa porque pone a disposición datos para comprender las problemáticas de la violencia de género a partir de estos casos que llegan a instancias judiciales. La recopilación de este tipo de datos puede ayudar a identificar patrones y tendencias que contribuyan a diseñar políticas públicas orientadas a la prevención de la violencia de género.

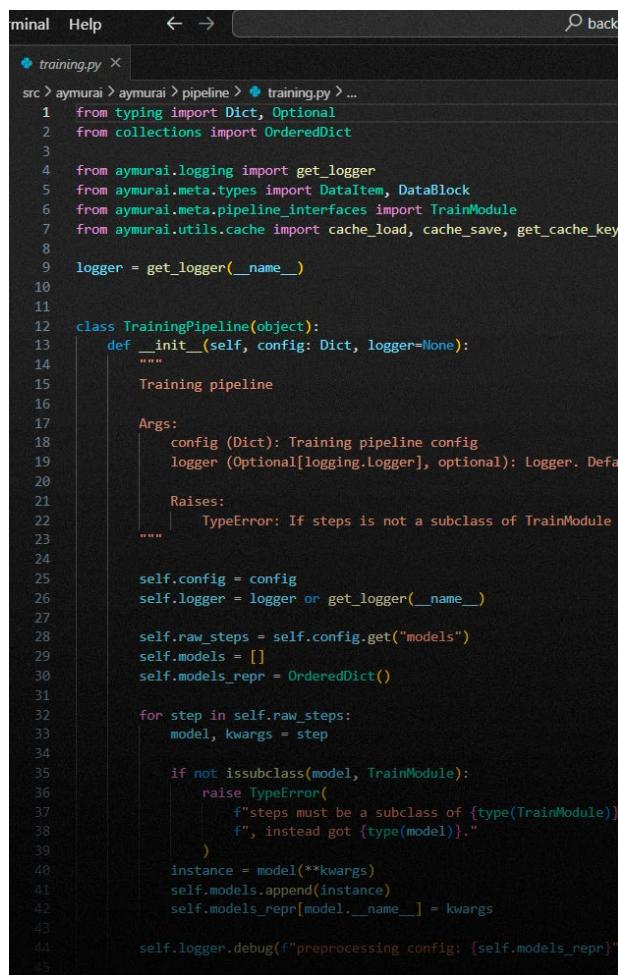
A partir del uso de AymurAI, el Juzgado publicó 6569 resoluciones judiciales.

El 30% de esas resoluciones se dictaron en casos vinculados a la violencia de género. En el 54% de esos casos estuvo presente la violencia física, en el 78% de esos casos el agresor era pareja, ex pareja o familiar de la denunciante, y en el 64% la violencia se ejercía de manera habitual.

Asimismo, el proyecto complementa la información brindada por la Oficina de Violencia Doméstica de la Corte Suprema de Justicia y los informes del Registro Único de Casos de Violencia de Género.

Por qué importa

- Había un gran volumen de datos no digitalizados. Esto, entre otras cuestiones, ralentizaba los procesos judiciales y volvía más opaca la Justicia.
- Permite diseñar políticas efectivas y dirigir los esfuerzos del Estado de manera más eficiente en la prevención de la violencia de género, a partir de la comprensión de la evolución y distribución de este fenómeno. Esto facilita el diseño, seguimiento y evaluación de políticas efectivas. De este modo, posibilita la evaluación de las acciones implementadas para combatir la violencia de género, contribuyendo a mejorar las intervenciones en esta área sensible.
- Este proyecto se enmarca dentro de las iniciativas de Justicia Abierta, que busca acercar la relación tradicionalmente distante entre el Poder Judicial y la ciudadanía, para así generar confianza y credibilidad en su labor.



```

minal Help ← → ⌂ back
src > aymurai > aymurai > pipeline > ● training.py > ...
1  from typing import Dict, Optional
2  from collections import OrderedDict
3
4  from aymurai.logging import get_logger
5  from aymurai.meta.types import DataItem, DataBlock
6  from aymurai.meta.pipeline_interfaces import TrainModule
7  from aymurai.utils.cache import cache_load, cache_save, get_cache_key
8
9  logger = get_logger(__name__)
10
11
12 class TrainingPipeline(object):
13     def __init__(self, config: Dict, logger=None):
14         """
15             Training pipeline
16
17             Args:
18                 config (Dict): Training pipeline config
19                 logger (Optional[logging.Logger], optional): Logger. Defa
20
21             Raises:
22                 | TypeError: If steps is not a subclass of TrainModule
23
24
25             self.config = config
26             self.logger = logger or get_logger(__name__)
27
28             self.raw_steps = self.config.get("models")
29             self.models = []
30             self.models_repr = OrderedDict()
31
32             for step in self.raw_steps:
33                 model, kwargs = step
34
35                 if not issubclass(model, TrainModule):
36                     raise TypeError(
37                         f"steps must be a subclass of {type(TrainModule)}"
38                         f", instead got {type(model)}."
39                     )
40
41                     instance = model(**kwargs)
42                     self.models.append(instance)
43                     self.models_repr[model.__name__] = kwargs
44
45                     self.logger.debug(f"preprocessing config: {self.models_repr}")
46
47
48
49
50
51
52
53
54
55
56
57
58
59
59
60
61
62
63
64
65
66
67
68
69
69
70
71
72
73
74
75
76
77
78
79
79
80
81
82
83
84
85
86
87
88
89
89
90
91
92
93
94
95
96
97
98
99
99
100
101
102
103
104
105
106
107
108
109
109
110
111
112
113
114
115
116
117
118
119
119
120
121
122
123
124
125
126
127
128
129
129
130
131
132
133
134
135
136
137
138
139
139
140
141
142
143
144
145
146
147
148
149
149
150
151
152
153
154
155
156
157
158
159
159
160
161
162
163
164
165
166
167
168
169
169
170
171
172
173
174
175
176
177
178
179
179
180
181
182
183
184
185
186
187
188
189
189
190
191
192
193
194
195
196
197
198
199
199
200
201
202
203
204
205
206
207
208
209
209
210
211
212
213
214
215
216
217
217
218
219
219
220
221
222
223
224
225
226
227
227
228
229
229
230
231
232
233
234
235
236
237
237
238
239
239
240
241
242
243
244
245
246
247
247
248
249
249
250
251
252
253
254
255
256
257
257
258
259
259
260
261
262
263
264
265
266
267
267
268
269
269
270
271
272
273
274
275
276
277
277
278
279
279
280
281
282
283
284
285
286
287
287
288
289
289
290
291
292
293
294
295
296
297
297
298
299
299
300
301
302
303
304
305
306
307
307
308
309
309
310
311
312
313
314
315
315
316
317
317
318
319
319
320
321
322
323
324
325
325
326
327
327
328
329
329
330
331
332
333
334
335
335
336
337
337
338
339
339
340
341
342
343
344
345
345
346
347
347
348
349
349
350
351
352
353
354
355
355
356
357
357
358
359
359
360
361
362
363
364
365
365
366
367
367
368
369
369
370
371
372
373
374
375
375
376
377
377
378
379
379
380
381
382
383
384
385
385
386
387
387
388
389
389
390
391
392
393
394
395
395
396
397
397
398
399
399
400
401
402
403
404
404
405
406
406
407
408
408
409
409
410
411
412
413
413
414
415
415
416
417
417
418
419
419
420
421
422
423
423
424
425
425
426
427
427
428
429
429
430
431
432
433
433
434
435
435
436
437
437
438
439
439
440
441
442
443
443
444
445
445
446
447
447
448
449
449
450
451
452
453
453
454
455
455
456
457
457
458
459
459
460
461
461
462
463
463
464
465
465
466
467
467
468
469
469
470
471
471
472
473
473
474
475
475
476
477
477
478
479
479
480
481
481
482
483
483
484
485
485
486
487
487
488
489
489
490
491
491
492
493
493
494
495
495
496
497
497
498
499
499
500
501
501
502
503
503
504
505
505
506
507
507
508
509
509
510
511
511
512
513
513
514
515
515
516
517
517
518
519
519
520
521
521
522
523
523
524
525
525
526
527
527
528
529
529
530
531
531
532
533
533
534
535
535
536
537
537
538
539
539
540
541
541
542
543
543
544
545
545
546
547
547
548
549
549
550
551
551
552
553
553
554
555
555
556
557
557
558
559
559
560
561
561
562
563
563
564
565
565
566
567
567
568
569
569
570
571
571
572
573
573
574
575
575
576
577
577
578
579
579
580
581
581
582
583
583
584
585
585
586
587
587
588
589
589
590
591
591
592
593
593
594
595
595
596
597
597
598
599
599
600
601
601
602
603
603
604
605
605
606
607
607
608
609
609
610
611
611
612
613
613
614
615
615
616
617
617
618
619
619
620
621
621
622
623
623
624
625
625
626
627
627
628
629
629
630
631
631
632
633
633
634
635
635
636
637
637
638
639
639
640
641
641
642
643
643
644
645
645
646
647
647
648
649
649
650
651
651
652
653
653
654
655
655
656
657
657
658
659
659
660
661
661
662
663
663
664
665
665
666
667
667
668
669
669
670
671
671
672
673
673
674
675
675
676
677
677
678
679
679
680
681
681
682
683
683
684
685
685
686
687
687
688
689
689
690
691
691
692
693
693
694
695
695
696
697
697
698
699
699
700
701
701
702
703
703
704
705
705
706
707
707
708
709
709
710
711
711
712
713
713
714
715
715
716
717
717
718
719
719
720
721
721
722
723
723
724
725
725
726
727
727
728
729
729
730
731
731
732
733
733
734
735
735
736
737
737
738
739
739
740
741
741
742
743
743
744
745
745
746
747
747
748
749
749
750
751
751
752
753
753
754
755
755
756
757
757
758
759
759
760
761
761
762
763
763
764
765
765
766
767
767
768
769
769
770
771
771
772
773
773
774
775
775
776
777
777
778
779
779
780
781
781
782
783
783
784
785
785
786
787
787
788
789
789
790
791
791
792
793
793
794
795
795
796
797
797
798
799
799
800
801
801
802
803
803
804
805
805
806
807
807
808
809
809
810
811
811
812
813
813
814
815
815
816
817
817
818
819
819
820
821
821
822
823
823
824
825
825
826
827
827
828
829
829
830
831
831
832
833
833
834
835
835
836
837
837
838
839
839
840
841
841
842
843
843
844
845
845
846
847
847
848
849
849
850
851
851
852
853
853
854
855
855
856
857
857
858
859
859
860
861
861
862
863
863
864
865
865
866
867
867
868
869
869
870
871
871
872
873
873
874
875
875
876
877
877
878
879
879
880
881
881
882
883
883
884
885
885
886
887
887
888
889
889
890
891
891
892
893
893
894
895
895
896
897
897
898
899
899
900
901
901
902
903
903
904
905
905
906
907
907
908
909
909
910
911
911
912
913
913
914
915
915
916
917
917
918
919
919
920
921
921
922
923
923
924
925
925
926
927
927
928
929
929
930
931
931
932
933
933
934
935
935
936
937
937
938
939
939
940
941
941
942
943
943
944
945
945
946
947
947
948
949
949
950
951
951
952
953
953
954
955
955
956
957
957
958
959
959
960
961
961
962
963
963
964
965
965
966
967
967
968
969
969
970
971
971
972
973
973
974
975
975
976
977
977
978
979
979
980
981
981
982
983
983
984
985
985
986
987
987
988
989
989
990
991
991
992
993
993
994
995
995
996
997
997
998
999
999
1000

```

Clasificación y apertura de datos de sentencias judiciales vinculadas a casos de violencia de género

Cómo funciona

[AymurAI](#)³ es un software que, a través de tecnologías de inteligencia artificial, identifica la información relevante en sentencias judiciales y genera conjuntos de datos abiertos sobre violencia de género. Es una herramienta “semi automática” porque no es independiente de la intervención y decisión humana, fundamental para lograr resultados seguros y acertados.

El sistema permite:

- Extraer la información relevante y necesaria para el análisis posterior. En los casos de violencia de género, los campos requeridos pueden ser: tipo de violencia, lugar, género, vínculo con el agresor, qué resolvió el juez o jueza en ese caso en particular, entre otros. El software utiliza modelos algorítmicos que combinan técnicas de reconocimiento de entidades nombradas (NER) para detectar y extraer los campos requeridos en las sentencias.
- Garantizar resultados seguros y acertados luego de la etapa de validación humana.
- Acceder a la información detectada agregada al conjunto de datos abiertos, luego de los mencionados procesos.

³ La herramienta AymurAI fue desarrollada por el Observatorio de Datos con Perspectiva de Género que forma parte de Data Género, una ONG feminista dedicada a investigar y repensar críticamente los procesos de trabajo con datos e inteligencia artificial.

De dónde se obtienen los datos

El proyecto recopila y clasifica los datos de sentencias judiciales digitalizadas vía AymurIA mediante tecnología de escáner, permitiendo la visualización digital de los expedientes. Este proceso de digitalización coexiste e interopera con el sistema oficial de gestión de expedientes judiciales, facilitando la gestión eficiente de expedientes dentro del juzgado. En paralelo, el juzgado trabaja en la apertura de datos judiciales, comenzando con la apertura de sentencias como datos no estructurados y evolucionando hacia la creación de una base de datos de sentencias en formato Excel.

Consideraciones finales

- Esta iniciativa no se limita simplemente a recopilar, clasificar y publicar datos, sino que también involucra la capacitación de los equipos que trabajan en el juzgado mejorando las capacidades estatales de la justicia en la gestión y análisis de datos.
- El proyecto AymurAI es una herramienta adaptable y de fácil implementación, gracias a su naturaleza de código abierto y a su flexibilidad en el desarrollo. Esto permite que pueda ser aplicado en diversas áreas de la justicia, y no se limita exclusivamente a casos de violencia de género, pudiéndose adaptar a diferentes problemáticas judiciales.
- Es fundamental impulsar y difundir el uso de este tipo de herramientas tecnológicas en otros juzgados y en otras provincias para optimizar la clasificación de datos, mejorar la transparencia y la rendición de cuentas en las decisiones judiciales, y capacitar al personal judicial en el uso de herramientas tecnológicas que contribuyan a la modernización del sistema judicial.

Seguir leyendo

- Para acceder al repositorio abierto de AymurIA visitar su repositorio en [Github](#).
- Para conocer más sobre el funcionamiento de AymurIA, leer las explicaciones de [este documento](#) de DataGénero.
- Para obtener información actualizada sobre las diversas etapas de un proceso judicial, que incluyen audiencias, sentencias y otros eventos relevantes, el Juzgado Penal, Contravencional y de Faltas N° 10 de la Ciudad de Buenos Aires mantiene una [cuenta de X \(Twitter\)](#) activa.
- Para acceder al conjunto de datos con perspectiva de género, compuesto por todas las decisiones desde agosto de 2016 hasta hoy, que tienen lugar en los casos que ingresan al [Juzgado N° 10](#), se encuentra abierta [esta planilla](#).
- Para obtener información detallada sobre la interpretación y el uso de estos datos consultar el [diccionario de set de datos con perspectiva de género](#).





Segmentación de la población objetivo en el diseño de políticas culturales



Las contribuciones a la redacción del caso se basaron en entrevistas con Julia Houllé del Ministerio de Cultura de la Nación - Dirección Nacional de Planificación y Seguimiento de Gestión.



Segmentación de la población objetivo en el diseño de políticas culturales



La plataforma del Registro Federal de Cultura (RFC) unifica el acceso de los ciudadanos a las convocatorias del Ministerio de Cultura de la Nación. La base de datos del RFC, asistida por algoritmos de aprendizaje automático, permite realizar un análisis segmentado de la población objetivo en grupos con características distintas y que indican la presencia de diferentes perfiles. A partir de este registro administrativo con datos de quienes desarrollan actividades culturales en el país se pudo segmentar mejor a la población destinataria de políticas culturales. La segmentación del RFC es el resultado de una colaboración entre el Área de Datos de Fundar y el Ministerio de Cultura de la Nación.

Antes de 2021, existían múltiples vías de acceso para postularse a las convocatorias del Ministerio de Cultura de la Nación. Había también una gran diversidad de circuitos y criterios que generaba una fragmentación de la información hacia el interior del organismo, así como también dificultades en la respuesta a las solicitudes de la ciudadanía. Ese año, se creó el RFC, una plataforma que unifica el acceso a las convocatorias (junto con los formularios de postulación) e integra la información con datos de otras fuentes. En abril de 2024 el RFC contaba con más de 150.000 personas inscriptas. Esta nueva plataforma colabora en la segmentación de la población destinataria de políticas culturales, lo cual contribuye con un mejor direccionamiento de las mismas basado en datos.

Durante la pandemia era esencial conocer los perfiles de los trabajadores de la cultura y la medida en que la pandemia los había dejado en situación de vulnerabilidad.

En ese momento, el RFC estaba conformado en un 75% por personas que se habían inscripto para postularse al [Apoyo Extraordinario Cultura Solidaria \(Resolución 210/2021\)](#). Este programa estaba dirigido a trabajadores y trabajadoras culturales afectados por las medidas de aislamiento.

Por qué importa

- El Ministerio de Cultura de la Nación carecía de bases de datos integradas sobre la gestión para conocer a quiénes se había alcanzado con sus políticas. La falta de una plataforma unificada limitaba significativamente la capacidad de análisis, segmentación y seguimiento de las políticas culturales.
- El RFC aporta información sobre trabajadoras y trabajadores del sector cultural y de esta forma contribuye al análisis de la cultura como actividad económica.
- El sistema también facilita el acceso a la información pública y la población también puede realizar su propio análisis independiente de los datos crudos.



Cómo funciona

La plataforma funciona con un esquema centralizado de bases de datos que permite hacer análisis e informes de manera ágil. La manera de agrupar a las personas inscriptas es mediante el análisis de *clustering*, una técnica de aprendizaje no supervisado. La utilización de esta metodología permite dividir a la población en grupos (*clusters*) con características distintas y que indican la presencia de diferentes perfiles.

Llevar a cabo una clusterización sobre la base de datos con información de personas permite:

- Descubrir patrones en los datos para agruparlos en segmentos (*clusters*) conformados por objetos (en este caso trabajadores de la cultura) similares.
- Encontrar grupos y perfiles distintos basados en las características de cada trabajador que esté inscripto en el RFC.
- Analizar y conocer en mayor profundidad a la población objetivo de la política pública.
- Evaluar el impacto de una política entre distintos subgrupos.

De dónde se obtienen los datos

Durante el año 2020, el Área de Monitoreo y Evaluación de la Dirección de Planificación y Seguimiento de Gestión comenzó a sistematizar las políticas del Gobierno Nacional que impactaban en el sector cultural. Se fortaleció la colaboración con dependencias de otros Ministerios y se allanó el camino para integrar a la base del RFC otros registros administrativos del Ministerio de Cultura y sus Organismos Descentralizados. Se trabajó en conjunto con áreas de estadística del Instituto Nacional del Teatro (INT), el Fondo Nacional de las Artes (FNA) y la Comisión Nacional de Bibliotecas Populares (CONABIP), entre otros. Además, se integraron bases de datos históricas y actuales para evaluar programas de los cuales no existían registros completos de beneficiarios.

Consideraciones finales

- La segmentación en diversos perfiles puede colaborar en ayudar a dar mayor precisión a los objetivos de cualquier política pública.
- El proyecto continúa en funcionamiento y se expandió para integrar registros administrativos de otros organismos y ministerios, lo que fortaleció la colaboración interinstitucional y mejoró la generación de información para la toma de decisiones en el ámbito cultural.
- En este sentido, también existen otras herramientas como el Sistema de Información Cultural de la Argentina (SInCA) que lleva adelante la Encuesta Nacional de Consumos Culturales (ENCC) y la Cuenta Satélite de Cultura (CSC), que pueden enriquecer este análisis con datos sobre consumo y producción del sector cultural.

Seguir leyendo

- Para mayor detalle sobre el Registro Federal de Cultura (RFC) dirigirse a su [sitio web oficial](#).
- Para más detalles sobre el proyecto, se puede explorar: el sitio web [Cultura en Datos](#). Allí se ofrece acceso a diversas bases de datos, informes de gestión, documentos metodológicos y guías de uso. Además, proporciona estadísticas detalladas sobre programas y convocatorias, indicadores de impacto y alcance de políticas culturales, así como información sobre programas específicos y la distribución geográfica de beneficiarios a través del [Mapa del Registro Federal de Cultura](#).
- Para acceder a bases de datos actualizadas consultar el [Portal de datos abiertos](#).
- Para conocer sobre la colaboración entre el [Área de Datos de Fundar](#) y la Dirección de Planificación y Seguimiento de Gestión del Ministerio de Cultura leer esta [Guía práctica](#) para caracterizar a la población objetivo de una política pública a partir de registros administrativos ([Avenburg et al. 2022](#)). Fue elaborada con el objetivo de identificar herramientas útiles para esta tarea dentro del ámbito de las políticas culturales.



Análisis de comportamiento turístico a partir de *big data*



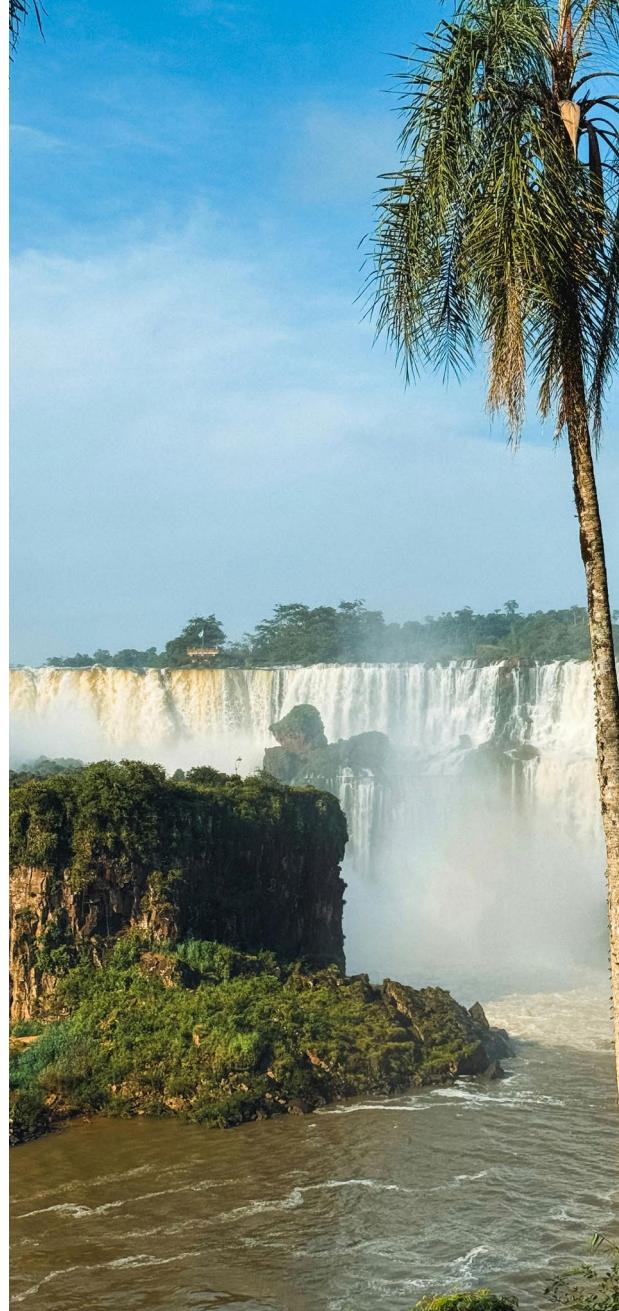
Las contribuciones a la redacción del caso se basaron en entrevistas con Elián Soutullo y otros actores involucrados en el proceso desde el Ministerio de Turismo de la Nación, así como con Rodrigo Castro de UBA-IC.

Análisis de comportamiento turístico a partir de *big data*

A partir de una base de grandes volúmenes de datos provenientes de señales de dispositivos móviles, se desarrolló una estrategia para generar información inédita sobre el comportamiento turístico en Argentina. Esta fuente de información novedosa permitió abordar preguntas de difícil respuesta con las fuentes de datos tradicionales para caracterizar al turismo en el país.

El desarrollo y uso de fuentes alternativas de *big data* permite explorar opciones para sortear algunas de las limitaciones existentes en las fuentes tradicionales. La Dirección Nacional de Mercados y Estadísticas (DNMyE) del Ministerio de Turismo y Deportes de la Nación dispone desde hace tiempo de diversas herramientas de información como los operativos estadísticos Encuesta de Viajes y Turismo de los Hogares (EVyTH), Encuesta de Turismo Internacional (ETI) y la Encuesta de Ocupación Hotelera (EOH) y registros administrativos como cruces migratorios, visitaciones a parques nacionales o aerocomercial.

Más cerca en el tiempo, también complementó estas encuestas con el desarrollo del Sistema de Información Turística de la Argentina (SINTA), que amplió la explotación de fuentes secundarias (registros administrativos de transporte terrestre o programas de asistencia y promoción) con el objetivo de fortalecer las herramientas de análisis del sector. Esta iniciativa nace de una época particularmente anómala para el sector, producto de la pandemia de COVID-19. Las restricciones a la movilidad de las personas significaron caídas drásticas en prácticamente todos los indicadores: retrocedió un 54% en 2020 y 33% en 2021 respecto de 2019. El turismo receptivo recibió un impacto aún más profundo: en 2020 cayeron 72% los viajes de turistas no residentes respecto del año anterior, mientras que en 2021 su nivel fue prácticamente nulo ([DNMYE, 2023](#)). De este modo, dado el nivel de granularidad que permiten las fuentes alternativas de *big data* como esta, su uso complementa y potencia la información proveniente de las fuentes de datos tradicionales.



La evolución y crecimiento de las fuentes de información relativas al turismo encontró en la exploración de nuevas fuentes alternativas la posibilidad de dar respuestas antes inabordables y permite caracterizar con mayor detalle dimensiones relevantes tales como el origen y el destino de viajeros.

Por qué importa

- El **SINTA**, más allá de contar con un amplio abanico de recursos de información para el estudio del sector turístico, encontraba límites para dar respuestas para el conjunto de la población objetivo.
- La Encuesta de Viajes y Turismo de los Hogares (EVyTH), el principal operativo para caracterizar el turismo interno en Argentina, proporciona únicamente información agregada a nivel de regiones turísticas con un enfoque trimestral y está diseñada para caracterizar el origen de los visitantes de los grandes aglomerados urbanos, excluyendo a residentes de ciudades medianas y pequeñas. En contraste, la base de grandes volúmenes de datos utilizada en este proyecto proporciona información de alta granularidad temporal y geográfica.
- El proyecto amplió la capacidad de análisis, planificación y desarrollo de la política turística. En particular, esta herramienta posibilita indagar con mucha mayor granularidad temporal y espacial sobre aspectos del comportamiento turístico de las personas.



Cómo funciona

La base recopila registros anónimos de dispositivos móviles, identificados mediante un código único destinado a publicidad o IFA (*Identifier for Advertising*), con georreferenciación a diario para los 528 departamentos de Argentina (a excepción de la Antártida e islas del Atlántico Sur), para el período de abril de 2019 hasta marzo de 2020. Entre otras cuestiones: (i) se pudo asignar residencia a 12.9 millones de los 17 millones de registros únicos, identificando que 2.8 millones realizaron viajes (origen) y (b) se identificaron más de 41 millones de registros de movimientos diarios de estos viajeros para todo el período de estudio. De este modo, se pudo caracterizar con mayor detalle los viajes analizados a partir de la mayor granularidad de la información espacial y temporal con la que se trabajó en el proyecto, dos limitaciones típicas de las fuentes tradicionales.

La incorporación de esta base de datos permite:

- Determinar el lugar de residencia para cada IFA (o usuario único) basado en la ubicación más frecuente durante el horario nocturno (*Common Evening Location - CEL*).
- Asociar cada IFA con las características socioeconómicas correspondientes a esa zona de residencia, vinculando las observaciones con información censal.
- Indagar sobre cortes temporales específicos.

De dónde se obtienen

los datos

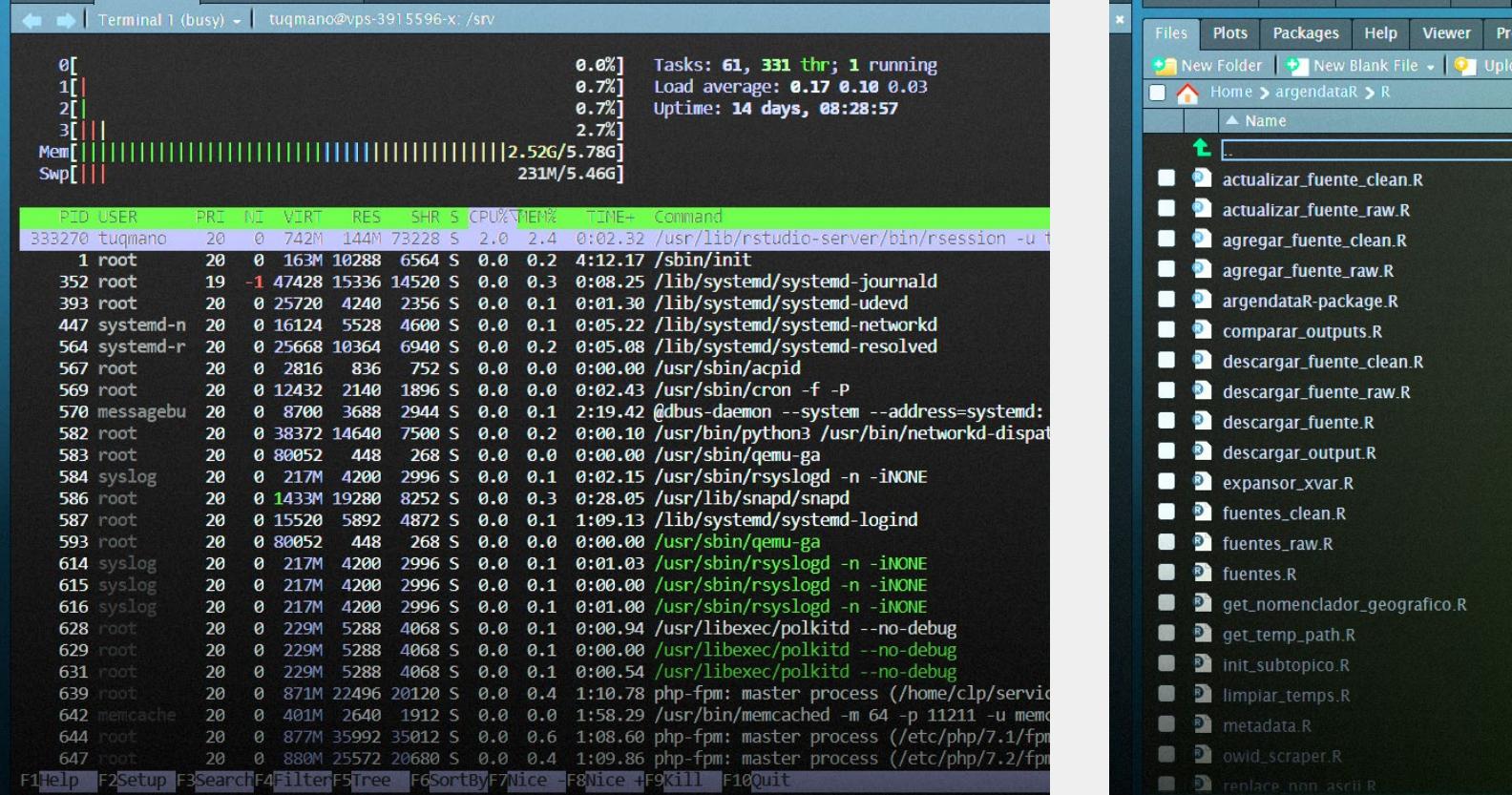
El Instituto Nacional de Promoción Turística (INPROTUR) adquirió una base de datos con información recolectada de dispositivos móviles, proveniente del registro de señales a través de aplicaciones instaladas por los usuarios y asociadas a un código de identificación único (IFA) utilizado para marketing. Esta base de datos, inicialmente utilizada para la segmentación de público objetivo en acciones de promoción digital, fue reutilizada en este proyecto para entender el comportamiento turístico. Además, se le añadió información socioeconómica del Índice de Vulnerabilidad Sanitaria ([Olego et. al, 2019](#)), a partir de la ubicación geográfica definida para cada usuario único. Lo anterior, combinado con fuentes tradicionales de uso del sector, como los operativos estadísticos y registros administrativos antes mencionados, permitió abordar aspectos difíciles de lograr con estrategias tradicionales.

Consideraciones finales

- Esta experiencia resalta la importancia de buscar y aprovechar los recursos existentes en el sector público, como los equipos especializados en universidades estatales, para abordar problemas específicos de manera efectiva y eficiente.
- La experiencia mostró la posibilidad de utilizar datos de manera creativa y reutilizar información que no fue inicialmente diseñada para los propósitos actuales. Además, se exploraron perfiles de segmentos del turismo argentino y se promovió el turismo en otras regiones. Se observó un genuino interés por parte de las provincias y las agencias de viaje, lo que indica que la herramienta tuvo un efecto positivo.

Seguir leyendo

- Para conocer más sobre el SINTA, visitar su [sitio web oficial](#).
- Para encontrar más información sobre cómo potenciar el turismo con el uso de *big data* se puede consultar la lectura de [De Ushuaia a La Quiaca: byte por byte. Cómo potenciar el sector turismo con el uso del big data \(Yankelevich et. al 2023\)](#). Este documento es el resultado del trabajo conjunto entre la DNMyE, el Laboratorio de Simulación de Eventos Discretos (SEDLab) del Instituto de Ciencias de la Computación ([ICC UBA-CONICET](#)) en la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires (FCExN-UBA), y el [equipo de Datos de Fundar](#).
- Para complementar la metodología presentada en el documento anterior, [en la entrada web de la Bitácora](#) de la DNMyE se presenta un ejemplo detallado que sigue una metodología similar. Este caso analiza datos sobre el turismo internacional receptivo, ofreciendo una perspectiva que enriquece la comprensión de los mercados turísticos y su potencialidad de expansión.



Conclusión

En este documento se presentaron experiencias de trabajo innovadoras de dependencias estatales que incorporaron la inteligencia artificial y fuentes de datos no tradicionales. Se trata de casos novedosos que dan cuenta de la importancia creciente de los datos en Argentina como herramienta clave para promover la transparencia y facilitar la toma de decisiones fundamentadas en políticas públicas.

La falta de continuidad en las políticas públicas también afecta a las áreas de datos.

Aunque se reconoce la importancia de una política de datos estable y sostenida en el tiempo, la falta de continuidad en las políticas públicas también afecta a las áreas de datos. La interrupción en la implementación de políticas de gestión de los datos puede crear vacíos de información y dificultar la planificación a largo plazo.

El éxito de estas experiencias se valora por su contribución a la transparencia en el acceso a la información pública, su utilidad de soporte para la gestión, la aceleración y ahorro de tiempo y recursos que propician en las respuestas del Estado a

certas problemáticas, la disponibilidad oportuna de información que brindan, el facilitamiento a la hora de realizar análisis segmentados de la población objetivo de políticas públicas, la clasificación de datos no estructurados y la ampliación de la capacidad de análisis, planificación y desarrollo de la política con mucha mayor granularidad temporal y espacial.

A su vez, estos casos conviven con varios desafíos. La interoperabilidad entre órganos de gobierno nuclea algunos de ellos, ya que para lograrla aún debe sobreponerse a: la fragmentación de bases de datos, la heterogeneidad entre dependencias estatales en la perspectiva sobre los datos, la diversidad de sistemas informáticos o la falta de un lenguaje común.

Necesitamos diseñar una estructura de gestión eficaz que coordine la implementación de una estrategia de gobernanza de datos. Para ello debemos optimizar la regulación para evitar la duplicación de normas y procedimientos. Asimismo, es crucial desarrollar y conservar el talento humano en el Estado, lo cual comienza por identificar y mapear las habilidades existentes. La mejora en el uso de datos para la planificación y toma de decisiones contribuirá a generar credibilidad en la información generada y reconstruir la confianza con la ciudadanía.

Bibliografía



- Avenburg, A., Houllé, J., Luvini, P. y Rodrigues Pires, M. (2022). [Guía práctica para caracterizar a la población objetivo de una política pública a partir de registros administrativos.](#)
- Ávila, R y Ruiz Nicolini, J. P. (2024). [Digital Public Infrastructure for Electoral Processes, an opportunity to increase trust and enhance democratic governance.](#) T20 Policy Brief.
- Bercovich, S., Szulmajster, I., Feldfeber, M. y Quiroga, Y. B. (2021). [Datos con perspectiva de género y justicia abierta: la experiencia del Juzgado 10.](#)
- Berniell, L., Hatrixk, A., Llambí, C., Durán, P., Olivera, M., Ontivero, L. J., y Ortega Grebenc, P. (2023). [Alertas tempranas para prevenir el abandono escolar: el caso de la Provincia de Mendoza.](#) CAF.
- DataGénero - Observatorio (2023). [AymurAI: Inteligencia Artificial responsable, para una justicia abierta y con perspectiva de género.](#)
- Delprato, M., Perusia, J. C. y Paparella, C. (2023). [Diagnóstico del abandono escolar en la educación secundaria en la provincia de Mendoza.](#) CIPPEC.
- Dirección de Planificación y Seguimiento de Gestión (s.f.). [Cultura en Datos.](#) Ministerio de Cultura de la Nación.
- Dirección de Planificación y Seguimiento de Gestión (s.f.). [Mapa Registro Federal de Cultura.](#) Ministerio de Cultura de la Nación.
- Dirección de Planificación y Seguimiento de Gestión (s.f.). [Registro Federal de Cultura.](#) Ministerio de Cultura de la Nación.
- Dirección General de Escuelas. (s.f.). [Sistema de Alerta Temprana de Trayectorias en Riesgo.](#) Ministerio de Educación, Cultura, Infancias y DGE.
- Dirección Nacional de Mercados y Estadística (2023). [Anuario Estadístico de Turismo 2021.](#) Ministerio de Turismo y Deportes.
- Dirección Nacional de Población. (s.f.). [Centro Nacional de Estudios de Población.](#) Ministerio del Interior.
- Dirección Nacional de Población. (s.f.). [Infraestructura de Datos Espaciales.](#) Ministerio del Interior.
- Dirección Nacional de Población. (s.f.). [Sistema Estadístico de Población.](#) Ministerio del Interior.
- Dirección Nacional Electoral (s.f.). [Sistema de Publicación de Resultados Electorales.](#) Ministerio del Interior, Obras Públicas y Vivienda.
- Luvini, P.; Dias, J. M.; Kunst, M.; Ruiz Nicolini, J. P. y Yankelevich, D. (2023). [Una estrategia de datos para la Administración Pública Nacional.](#)
- Ministerio de Educación de Guatemala (2019). [Guía para promover la transición exitosa de sexto primaria a primer grado del Ciclo Básico.](#) Ministerio de Educación de Guatemala.
- Ministerio del Interior, Dirección Nacional Electoral. (2024, marzo 11). [Disposición 1/2024.](#) Boletín Oficial de la República Argentina.
- Naciones Unidas y CEPAL (2013). [Consenso de Montevideo sobre población y desarrollo.](#) CELADE.
- Olego, T., Rosati, G. t Vázquez Brust, A. (2019). [Vulnerabilidad Sanitaria en Argentina: Documento de Trabajo 2019-01.](#) Fundación Bunge y Born.
- Quiroga, Y. y Mandolesi, M.A. (2020). [Apertura de Datos con Perspectiva de Género.](#) Pensamiento penal.
- Paparella, C. y Xanthopoulos J. (2023). [Metodología para el desarrollo de un SAT provincial y evaluación de las condiciones de implementación.](#) CIPPEC.
- Prokhorenkova, L., Gusev, L., Vorobev, A., Dorogush, A. V. y Gulin, A. (2018). [CatBoost: Unbiased boosting with categorical features.](#) Proc. NeurIPS, Integravke Differentials.
- RENAPER. (s.f.). [Dirección Nacional de Población.](#) Ministerio del Interior.
- Yankelevich, D., Soutullo, E., Juara, J. G., Ortiz de Zárate, J. M., Ruiz Nicolini, J. P., Kunst, M., Zapatero, M., Grimson, R. y Castro, R. (2023). [De Ushuaia a La Quiaca: byte por byte. ¿Cómo potenciar el sector turístico con big data?](#) Fundar.

Acerca del equipo autoral

Juan Pablo Ruiz Nicolini

Coordinador de Datos de Fundar

Licenciado en Ciencia Política y Gobierno y magíster en Ciencia Política por la Universidad Torcuato Di Tella. Cumplió funciones como Director Nacional de Mercados y Estadística en el Ministerio de Turismo y Deportes de la Nación, donde desarrolló el Sistema de Información Turística de la Argentina. Investigó temas relacionados con partidos políticos, elecciones y reforma electoral. Se desempeñó como docente de ciencia de datos en UTDT.

Mariana Kunst

Coordinadora de Datos de Fundar

Licenciada en Economía y magíster en Métodos Cuantitativos para la Gestión y Análisis de Datos por la Universidad de Buenos Aires. Se desempeñó como coordinadora del Sistema de Información Cultural de la Argentina (SInCA) y como asesora en programas vinculados a las industrias culturales y análisis de información en el Ministerio de Cultura de la Nación. Realizó tareas de investigación relacionadas con empresas y organizaciones del sector cultural desde la Historia Económica. Actualmente es docente en la Universidad de Buenos Aires.

Juan Manuel Dias

Científico de Datos de Fundar

Licenciado en Sociología por la UBA y maestrando en Estadística de la UNTREF. Es egresado de la carrera de ciencia de datos de la EANT y de la Diplomatura de Ciencias Sociales Computacionales de la UNSAM. Trabajó en investigaciones de mercado y de opinión pública en el sector privado y tiene una amplia experiencia en la administración pública, en las áreas de evaluación de políticas e innovación de procesos vinculados a la captación y análisis de información. Actualmente es docente de estadística en la UNPAZ.

Equipo Fundar

Dirección ejecutiva: Martín Reydó

Dirección de proyectos: Lucía Álvarez

Coordinación editorial: Juan Abadi

Revisión institucional: Marcelo Mangini

Corrección/Edición: Juan Abadi

Diseño: Micaela Nanni

Ruiz Nicolini, Juan Pablo
Usos inteligentes de datos en el Estado / Juan Pablo Ruiz Nicolini ; Mariana Kunst ;
Juan Manuel Dias. - 1a ed - Ciudad Autónoma de Buenos Aires : Fundar , 2024.
Libro digital, PDF

Archivo Digital: descarga y online
ISBN 978-631-6610-25-6

1. Bases de Datos. 2. Administración Pública. 3. Sistemas de Gestión de Bases de
Datos. I. Kunst, Mariana II. Dias, Juan Manuel III. Título
CDD 320.014

ISBN 978-631-6610-25-6





fundar

www.fund.ar

info@fund.ar