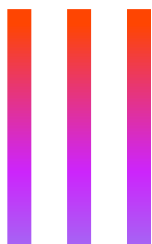


CHECKLIST

Innovar con Ciencia de Datos en el sector público



Innovar con Ciencia de Datos en el sector público

Autoría

María Vanina Martínez

Directora del Programa de Ciencia de Datos e Inteligencia Artificial (PCDIA) de la Fundación Sadosky. Doctora de Ciencias de la Computación (University of Maryland) e investigadora del CONICET.

Victoria Gisel Dumas

Consultora del PCDIA de la Fundación Sadosky. Doctora en Ciencias Químicas (Universidad de Buenos Aires) y profesora de Visualización y Comunicación en la Maestría en Ciencia de Datos de la Universidad de San Andrés.

Marianela Sarabia

Consultora del PCDIA de la Fundación Sadosky. Magister en Economía Aplicada para el Desarrollo (UniTo & SciencesPo), y profesora de Políticas Públicas y Machine Learning (Universidad Nacional de Guillermo Brown).

Ivana Feldfeber Kisilevsky

Consultora del PCDIA de la Fundación Sadosky. Posgrado en Ciencia de Datos, Aprendizaje Automático y sus Aplicaciones (Universidad Nacional de Córdoba) y Directora Ejecutiva del Observatorio de Datos con Perspectiva de Género "DataGénero".

Colaboración

Manuela Cerdeiro

Fundación Sadosky

Lucas Somacal

Fundación Sadosky

Inés Roggi

Fundación Sadosky

Daniel Yankelevich

Fundar

Mariano Cervelini

Fiscalía de la Provincia de Buenos Aires

Marta Kisilevsky

Ministerio de Educación de la Nación

Gabriela Reta

Universidad Nacional de General Sarmiento

María del Mar Monti

Ministerio de Defensa de la Nación

Edición

Laura Ramírez Rivillas

Fundación Sadosky

Alejandra Stafetta

Fundación Sadosky

Federico Rey

Fundación Sadosky

Diseño

Jaqueline Schaab

Fundación Sadosky

Facundo Manini

Fundación Sadosky

Autoridades de la Fundación

Dr. Manuel Sadosky

Daniel Filmus

Presidente

Fernando Schapachnik

Director Ejecutivo

Hoja de ruta: Innovar con datos en el sector público

/ Ivana Feldfeber Kisilevsky...

[et al.]. - 1a ed. - Ciudad Autónoma de Buenos

Aires : Fundación Sadosky, 2022.

Libro digital, PDF

Archivo Digital: descarga

ISBN 978-987-48926-0-7

1. Computación. 2. Políticas Públicas. 3.
Recopilación de Datos. I. Feldfeber Kisilevsky, Ivana.
CDD 352.75



Checklist para implementar Ciencia de Datos e Inteligencia Artificial en el sector público

Definición del problema o la necesidad

- ¿Qué quiero resolver o mejorar?
 - ¿Cuál es el alcance de esta problemática?
 - ¿Cuál es la población destinataria?
 - ¿Tengo en claro qué impactos genera? ¿Qué impactos quiero generar? ¿Hay algún aspecto negativo entre los posibles impactos?
 - ¿Qué es lo indispensable que debo comunicar o difundir en las distintas instancias?

Herramientas tecnológicas y el uso de CDIA

- ¿Se puede vincular esta problemática con las herramientas tecnológicas que quiero utilizar? ¿De qué manera?
 - ¿Sería algo novedoso? ¿O una mejora de algo existente?
 - ¿Cómo me ayudaría la herramienta ideal a mejorar la gestión, la política pública o el servicio?
 - ¿La herramienta ideal implica modificar algún proceso de trabajo o circuito de gestión? ¿Sobre qué/ quiénes impactaría en este último caso? ¿Cómo podríamos trabajar esta transición?
- ¿Tengo datos sobre la problemática? ¿Cuáles necesitaría sistematizar y/o recolectar?
- ¿La herramienta ideal utilizará CD, aprendizaje automático o la IA?
 - ¿Serían las mejores herramientas para ayudar a resolver este problema?
 - ¿Qué otras herramientas o abordajes podría utilizar?
 - ¿Es posible que esta herramienta sea usada de manera incorrecta y genere efectos no deseados? ¿Cuáles? ¿Son evitables?

- ¿Pienso en desarrollar o en adquirir y adaptar la herramienta de CDIA?
 - En caso de adquirirla, ¿existe documentación y acompañamiento para su implementación? ¿Quién del equipo la probaría o la adaptaría?

Equipo de trabajo

- ¿Mi equipo de trabajo es interdisciplinario? ¿Qué capacidades necesito reforzar?
- ¿Hay planificadas instancias de intercambio y trabajo en conjunto entre el equipo experto en la problemática y el equipo técnico?

Conocimiento en profundidad de los datos

- Si quiero implementar un modelo de aprendizaje automático o IA:
 - ¿Tengo los datos que requiere el algoritmo?
 - ¿Conozco en profundidad mi set de datos?
 - ¿Con qué información adicional me gustaría contar?
 - ¿Mis datos tienen calidad?
 - ¿Qué preguntas tengo que hacerme para comprender mis datos en profundidad?

Infraestructura

- ¿Qué infraestructura necesito para realizar y mantener una herramienta de CDIA?
 - ¿Dónde están almacenados los datos?
 - ¿Van a ser necesarios servidores externos para alojarlos o realizar operaciones sobre éstos?
- ¿Qué medidas de seguridad estamos planificando para que estos datos estén resguardados y no haya filtraciones?

- ¿Con qué equipamiento cuento y qué necesito? ¿Las computadoras actuales permiten leer los archivos y procesar los datos que necesito?

Documentación

- ¿Estoy planificando documentar mi herramienta para que sea replicable por otros equipos? ¿Y documentar las decisiones sobre los datos y el proceso?
- ¿Tengo pensado desarrollar un manual de uso de la misma para personas no técnicas? ¿Qué capacidades adicionales necesitamos para producir este tipo de manual (por ejemplo: expertise en didáctica, en redacción simple, etc.)?
- ¿Cómo y dónde voy a documentar el proyecto? ¿Hay una persona del equipo designada? ¿Es un proceso colaborativo?
- ¿Qué herramientas tengo a disposición para hacerlo?
- ¿Tengo en cuenta que la documentación tiene que ser interdisciplinaria y no solamente del equipo técnico?
 - Si otro equipo toma la herramienta que diseñé y la documentación: ¿puede replicarla o editarla fácilmente? ¿Qué acciones debemos tomar para que si el equipo cambia, el proyecto siga funcionando?

Impacto

- ¿Podemos prever el impacto desde el diseño de la herramienta?
- ¿Cómo podemos planificar una evaluación de impacto de la herramienta?
- ¿Está prevista la difusión o capacitación sobre la herramienta? ¿Cuenta con posibilidad de escalamiento?

Revisión y monitoreo continuo de la herramienta/proyecto

- ¿Qué instancias de revisión o monitoreo continuo estamos considerando para este proyecto?
- ¿Con cuánta periodicidad vamos a monitorear el desempeño de estas herramientas?
 - Si identifico que hay nuevos datos no contemplados en la fase de aprendizaje del modelo, ¿podemos pautar un análisis de los nuevos datos? ¿Tengo prevista la actualización del modelo en función de eso?

Algunas instancias de revisión:

- En el inicio del proyecto: ¿estamos haciéndolo de manera correcta?
- Durante el proyecto: ¿lo hemos diseñado bien?
- Después del proyecto: ¿sigue funcionando como queríamos que lo hiciera cuando lo diseñamos?¹

Sustentabilidad y ambiente

- ¿Conocemos el impacto ambiental de las herramientas tecnológicas que vamos a utilizar? ¿Y de los equipamientos e infraestructura?
- ¿Estamos teniendo en cuenta este impacto a la hora de elegir las herramientas a utilizar?

Normativa / leyes

- ¿Es la herramienta segura y se adecúa a las legislaciones vigentes en materia de seguridad y privacidad de la información?
- Si se trabaja con datos sensibles, ¿qué medidas de seguridad se están utilizando para protegerlos?
- ¿El uso de estos datos interfiere con los derechos de las personas cuyos datos están siendo utilizados? Si es así, ¿Hay alguna forma menos intrusiva de llegar al objetivo deseado?
- A partir de la herramienta ¿necesito generar algún registro formal?

¹ Véase www.gov.uk/government/publications/data-ethics-framework

Transparencia

- ¿Qué acciones tengo en cuenta para garantizar la transparencia de la herramienta?
- ¿Hay instancias planificadas para explicar el uso de los datos y los resultados de los algoritmos a la población?
- ¿Cómo se podrían co-diseñar las herramientas de CDIA con la sociedad civil?
- ¿Qué partes de los desarrollos de estas herramientas y de los datos se pueden hacer públicas para que cualquier persona tenga acceso?

Sesgos

- ¿Puedo identificar algún sesgo en mis datos?
- ¿Existen prejuicios en mi equipo con respecto a la necesidad o problema que queremos abordar? ¿Cuáles son?
- ¿Qué riesgos encontramos en la implementación de la herramienta o ante los resultados que brinda?
- ¿Qué acciones concretas puedo realizar para mitigar y balancear los sesgos de los datos que uso y del modelo? ¿Tengo previsto exponerlos y/o validarlos por fuera de mi propio equipo?



Innovar con
Ciencia de Datos
en el sector
público

Fundación
SADOSKY

