



Attribution ShareAlike (CC BY-SA) 

Hoja de trabajo para definir el alcance de un proyecto de ciencia de datos¹

1. **Nombre del proyecto:** Desarrollando modelos predictivos de riesgo para apoyar las decisiones de la línea directa de maltrato infantil en el Condado de Allegheny (Pennsylvania)²
2. **Nombre de la organización:** Departamento de Servicios Humanos (DHS) del Condado de Allegheny

1. Descripción del problema:

1.1 ¿Cuál es el problema al que te enfrentas?

El condado de Allegheny cuenta con 2 líneas de emergencia donde se puede denunciar maltrato o abuso infantil: una línea de emergencia del condado, y una línea de emergencia especializada, “Childline”. Las denuncias pueden clasificarse como CPS (servicios de protección infantil) y GPS (servicios de protección general). Cuando una denuncia se clasifica como CPS, se genera una investigación exhaustiva inmediatamente para ver si el niño/a y su entorno debe ser intervenido por el condado, y se considera que ese hogar ya es parte del sistema. Cuando una denuncia se clasifica como GPS el personal de la línea de emergencia del Condado debe decidir en alguna de las siguientes acciones: no investigar ni intervenir, investigar la denuncia de manera general o investigar a profundidad la denuncia. Se dice que un hogar entra en el sistema desde el punto en que se mandata una investigación exhaustiva.

De la totalidad de las denuncias que se clasifican como GPS, un 48% son

¹ Creado por el Center for Data Science and Public Policy de la Universidad de Chicago y el GobLab de la Universidad Adolfo Ibáñez. Esta descripción de proyecto fue creado por GobLab de la Universidad Adolfo Ibáñez, como parte de un currículum disponible [aquí](#). Puede utilizarlo citando a sus creadores y adaptarlo según la Licencia Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0), los términos y condiciones están disponibles [aquí](#)

² Esta hoja de trabajo ha sido creada por el GobLab UAI con fines pedagógicos. Es un ejercicio retrospectivo hecho a partir de información pública, que no necesariamente representa la formulación inicial del proyecto.





ingresadas al sistema (abril 2010-mayo 2016)

El problema es que de las denuncias que no terminaron en una intervención, en el 53% de los casos se volvía a denunciar en un plazo de 2 años. Por lo tanto, el sistema actual de decisión deja afuera a niños que requieren apoyo del condado. Dado el alto flujo de llamadas, y el tener que acceder a datos familiares de forma manual de otras fuentes, los operadores de las líneas de emergencia deben tomar decisiones en poco tiempo y no siempre alcanzan a realizar un perfil de riesgo adecuado del niño.

1.2 ¿Quién/qué se ve afectado por este problema? (Cierta tipo de personas, organizaciones, barrios, medio ambiente)

Niños que no recibieron la intervención adecuada de la autoridad. Familias que sufrieron un estigma negativo por haber sido intervenidas cuando no era realmente necesario o que no recibieron la intervención adecuada. Funcionarios del sistema de protección infantil del condado que debieron investigar casos de bajo riesgo, dejando de lado otros casos más riesgosos. Vecinos de Allegheny, cuyos impuestos no son usados de forma eficiente.

1.3 ¿Cuántas personas/organizaciones/localidades/etc. se ven afectadas y en qué medida? (P. ej., tiempo medio de espera para una cirugía, número de estudiantes que abandonan los estudios, costo debido a la evasión fiscal, etc.)

Actualmente el Censo estima que el condado de Allegheny cuenta con 1.218.452 habitantes, de los cuales un 18,7% son menores de 18 años (227.850) y 5,2% son menores de 5 años (63.359)³.

En el año 2017 se recibieron 15.678 llamados, de los cuales 11.751 fueron clasificados como GPS.

1.4 ¿Por qué resolver este problema es una prioridad para tu organización?

³ U.S Census Bureau. Estimaciones a 01 de Julio 2018





Gob_Lab UAI

UNIVERSIDAD ADOLFO IBÁÑEZ



El Departamento de Servicios Humanos (DHS) del Condado de Allegheny tiene como misión la creación de un sistema de servicios humanos accesible, culturalmente competente e integral que garantice servicios para la población del Condado de Allegheny, especialmente aquellos más vulnerables. En 2014 se inició un proceso que busca mejorar la capacidad de realizar decisiones basadas en datos de manera eficiente y consistente, asegurar que los recursos públicos se destinen equitativamente según vulnerabilidad, y mejorar aspectos generales de la salud y bienestar de los residentes del condado de Allegheny.





2. Objetivos (en orden de prioridad)

- La solución técnica (p. ej., modelo predictivo) no es el objetivo.
- El objetivo debe ser medible.
- Alcanzar el objetivo debería ayudar a solucionar el problema.
- Algunos objetivos habituales son mejorar/maximizar/aumentar o disminuir/mitigar/reducir un resultado o una métrica.
- Las limitaciones típicas incluyen el presupuesto, la falta de capital humano, las restricciones legales, la voluntad política y la licencia social.
- Considera las compensaciones entre objetivos contrapuestos.

	Objetivo	Limitaciones
1	Disminuir la cantidad de casos de niños que no son visitados tras una denuncia y que en los siguientes dos años, son separados de su familia tras una decisión judicial.	Sólo se visitan hogares que han sido denunciados.
2	Aumentar la proporción de casos visitados tras una denuncia, que sí requieren de intervención del condado.	No se pueden visitar todos los casos denunciados al condado, se visita aproximadamente un 50%.
3	Disminuir la proporción de casos visitados tras una denuncia, que no requieren de intervención del condado.	





3. Acciones

- Las acciones son lo que las instituciones pueden hacer respecto de un problema al asignar recursos, por ejemplo, inspeccionar instalaciones, brindar servicios preventivos, promoción, etc.
- Las acciones ideales deberían ayudarte a alcanzar el objetivo definido anteriormente.

	Acción 1	Acción 2	Acción 3
Acción: <i>P. ej., informar al propietario de un automóvil sobre cómo renovar su permiso</i>	Decidir si una denuncia debe ser sujeta a investigación.	Ingresar a un niño al sistema de protección infantil.	
¿Quién ejecuta la acción? <i>P. ej., departamento de TI (envía correo electrónico)</i>	Servicios Infantiles Allegheny (call-center)	Servicios Infantiles Allegheny	
¿Sobre quién/qué se está realizando la acción? <i>P. ej., propietario de automóvil</i>	Niños de hogares de los cuales se recibe denuncia telefónica y su entorno familiar.	Niños de hogares de los cuales se recibe denuncia telefónica y su entorno familiar.	
¿Con qué frecuencia se toma la decisión de realizar esta acción? <i>P. ej., anualmente</i>	Después de cada llamado telefónico.	Después de la denuncia telefónica (CPS)/después de la investigación (GPS).	





¿Qué canales se están usando/se pueden usar para realizar esta acción? P. ej., correo electrónico	Reuniones presenciales de los operadores y sus supervisores	Visitas presenciales al hogar denunciado.	
Otra información útil sobre la acción	En las reuniones presenciales se investigan otras fuentes de información que pueden ser útiles para tomar la decisión.	No existe un protocolo estándar para tomar	





4. Datos

- Los datos tienen que conectarse con las acciones que respaldan, de modo que la organización pueda alcanzar su objetivo.
- Los proyectos típicos de ciencia de datos usan datos administrativos como la fuente primaria de datos y la mejoran con fuentes de datos disponibles públicamente (censo, otros datos abiertos). La asociación con el sector privado u organizaciones sin fines de lucro podría ser una forma de obtener datos que podrían faltarte a nivel interno.

A. ¿Qué datos tienes internamente?

(agregar columnas de más fuentes si corresponde)

	Datos de los llamados ingresados al sistema	Fuente de datos 2	Fuente de datos 3
¿Qué contiene? P.ej., registros de admisión y alta de hospitales a nivel nacional	Información histórica de denuncias y demográficas de los hogares desde abril 2010 a abril 2014.		
¿Qué nivel de granularidad? P. ej., transacción, persona, organización, ubicación	A nivel de familia, miembros de la familia		





¿Con qué frecuencia se recopila/actualiza una vez que se capta?	En tiempo real.		
P. ej., en tiempo real, a diario, semanalmente, mensualmente, anualmente, de manera excepcional			
¿Tiene identificadores únicos y confiables que se pueden vincular con otras fuentes de datos? P. ej., RUN	Sí, identificador de los miembros del hogar en el momento de la denuncia.		
¿Quién es el propietario interno de los datos? P. ej., hospitales	Servicio Infantil Allegheny		
¿Cómo se almacena? P. ej., en una base de datos, PDF, Excel	Base de datos (Allegheny Datawarehouse)		
Comentarios adicionales			





B. ¿Qué datos puedes obtener de fuentes externas, privadas o públicas?

	Prisión del Condado	Libertad Condicional Juvenil	Servicio de bienestar público
¿Qué contiene?	Información de interacciones pasadas con la cárcel del condado.	Información histórica de relación con la cárcel juvenil de Allegheny.	Información de personas que han sido parte de alguno de los programas de bienestar (asistencia temporal, asistencia general, subsidio al ingreso, subsidio alimenticio, etc.)
¿Qué nivel de granularidad?	A nivel de persona.	A nivel de persona.	A nivel de persona.
¿Con qué frecuencia se recopila/actualiza una vez que se capta?	Mensualmente	Mensualmente	Mensualmente
¿Tiene identificador es únicos y confiables que se pueden vincular con otras fuentes de datos?	Sí, identificadores personales.	Sí, identificadores personales.	Sí, identificadores personales.





¿Quién es el propietario interno de los datos?	Prisión de Allegheny.	Oficina de Libertad Condicional Juvenil del Condado de Allegheny	Condado de Allegheny.
¿Cómo se almacena?	Excel/SQL	Excel/SQL	Excel/SQL
Comentarios adicionales			

	Programas de Salud del Comportamiento	Estimación de Pobreza de Censo
¿Qué contiene?	Información histórica de pacientes que han recibido ayuda de la oficina o que hayan sido diagnosticados.	Nivel de pobreza en cada código postal (un poco más que una manzana)
¿Qué nivel de granularidad?	A nivel de persona	A nivel de más de una manzana, menos que un barrio.
¿Con qué frecuencia se recopila/actualiza una vez que se capta?	Mensualmente	Cada 10 años.
¿Tiene identificadores únicos y confiables que se pueden vincular con otras fuentes de datos?	Sí, identificador personal.	Sí, código postal.





¿Quién es el propietario interno de los datos?	Oficina de Salud del Comportamiento del condado de Allegheny.	Gobierno Central.
¿Cómo se almacena?	Excel	Bases de datos
Comentarios adicionales	Esta base incluye: servicios de alcohol y drogas, servicios relacionados con poder judicial, servicios de salud mental, entre otros.	

C. **En un mundo ideal, ¿existen datos adicionales que te gustaría obtener/reunir que serían relevantes para este problema?** (Encuestas, circuito cerrado de televisión, registros telefónicos, ADN, diferente frecuencia o granularidad para datos disponibles actualmente, etc.)

5. Análisis

- Los proyectos típicos de ciencia de datos incluyen una combinación de análisis.
- El análisis no es el objetivo del proyecto.
- Elige el análisis adecuado para el problema correcto





- Debes probar el análisis, y el proceso de validación debe coincidir con tu objetivo.

	Análisis 1:	Análisis 2:	Análisis 3:
Tipo de análisis (descripción, predicción, detección, cambio de comportamiento)	Predicción		
Propósito del análisis (p. ej., entender el comportamiento histórico de las personas, estimar el riesgo de enfermedad de un paciente)	Predecir el riesgo maltrato de un niño, al recibir el llamado, para reforzar, o desestimar, la decisión de intervención en el hogar.		
¿Qué acción respaldará este análisis?	Acción 1: la decisión de investigación (ingresarlos al sistema)		
¿Cómo validarás este análisis usando datos existentes? (P. ej., usar datos históricos, ejecutar un ensayo aleatorio controlado)	Entrenamiento de modelo predictivo en base a datos históricos.		





6. Consideraciones éticas

<p>Privacidad</p> <p>¿Estás trabajando con datos personales y/o confidenciales identificables a nivel individual?</p>	<p>Sí, existen datos personales y sensibles, y debemos usar indicadores innominados.</p>
<p>Transparencia</p> <p>¿Qué partes interesadas deberían estar al tanto de qué partes del proyecto? (Las partes interesadas suelen incluir a formuladores de políticas, trabajadores de primera línea, personas que se verán afectadas por las acciones, etc.)</p>	<p>Gobierno regional de Allegheny, usuarios del sistema.</p>
<p>Discriminación/equidad</p> <p>¿Existen grupos específicos para los que deseas garantizar la equidad de los resultados?</p>	<p>Existe un sesgo racial presente en los datos históricos. Los niños afroamericanos tienen casi 3 veces más probabilidades de tener interacción con un servicio de protección infantil que niños blancos, debido a que la población tiende a denunciar más a las familias afroamericanas. También tienen más probabilidades de ser ingresados al sistema una vez se realiza el llamado. Se contemplará revisar el efecto que tiene el modelo en las disparidades del sistema.</p>
<p>Licencia social</p> <p>Si la población completa del país se entera de tu proyecto, ¿lo aprobará?</p>	<p>Hay que realizar un trabajo de difusión cuidadoso, siempre es complejo hablar de algoritmos en intervenciones con personas.</p>





<p>Rendición de cuentas ¿Quiénes son las personas responsables de todos los elementos anteriores?</p>	<p>Director del Departamento de Servicios Humanos Allegheny.</p>
<p>Otras consideraciones, como consentimiento, legales, etc.</p>	<p>Al trabajar con datos sensibles, es importante usar identificadores innominados para resguardar la privacidad de las personas.</p>

7. **¿Qué prueba de campo o ensayo aleatorio controlado puedes diseñar para validar el proyecto en terreno?** Los resultados que medirás deberían coincidir con tus objetivos.

Cruzar los datos de llamados con la información del Hospital de Pittsburgh, medir de esta forma si niños con alto riesgo estimado han tenido que acudir al sistema de urgencia. De esta forma podemos medir de manera empírica si un niño con un alto riesgo estimado, lo es también de riesgo real.

8. **¿Cuáles son las organizaciones externas y los departamentos internos que necesitarán estar involucrados?**

(Generalmente, los proyectos de ciencia de datos requieren la participación de propietarios de datos, propietarios de infraestructura de TI, propietario del problema, personas de analítica)

Organización/departamento	Descripción de la participación deseada	Nombre/rol de la contraparte
Oficina servicios humanos de Allegheny	Coordinar trabajo interno	Marc Cherna/Director





TI	Implementación, en tiempo real, del algoritmo en los computadores.	
Departamento de Estadística	Codificación de variables para el análisis. Actualización permanente	
Child Hotline	Capacitar a los operadores telefónicos en la nueva herramienta	
Comunidades de interés (e.g. representantes de comunidad afroamericana, académicos, etc.)	Brindar sus opiniones para ser consideradas en el diseño e implementación del sistema.	

Esta hoja de trabajo fue desarrollada originalmente por el Center for Data Science and Public Policy de la Universidad de Chicago. Para más información sobre nuestros programas y trabajo, por favor visita <http://datasciencepublicpolicy.org> o escríbenos a info@datascienceforsocialgood.org

Esta versión de la hoja de trabajo ha sido actualizada a través de una colaboración entre el GobLab UAI, Carnegie Mellon University e ITAM.

El GobLab UAI es el laboratorio de innovación de la Escuela de Gobierno de la Universidad Adolfo Ibáñez. Su misión es promover el uso de la ciencia de datos en el sector público con el fin de mejorar la gestión pública y tener más políticas públicas basadas en evidencia. Capacita a funcionarios públicos y realiza investigaciones y proyectos aplicados en asociación con organismos gubernamentales. Para obtener más información, visita <https://goblab.uai.cl> o envía un correo electrónico a goblab@uai.cl

