



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

1. Nombre del proyecto: Detectando riesgo de suicidio en aplicaciones de chat²

2. Nombre de la organización: Fundación Todo Mejora

3. Descripción del problema:

3.1 ¿Cuál es el problema al que te enfrentas?

La fundación Todo Mejora es una Organización no Gubernamental cuyo objetivo es promover el bienestar de niños, niñas y adolescentes que sufren bullying y comportamiento suicida, producto de ser discriminados en base a su orientación sexual, identidad y expresión de género. En Chile, el suicidio es la primera causa de muerte entre las personas de 15 a 24 años.³ En Chile, según datos de 2015, cada 2.8 días se suicida un menor de edad.⁴

Todo Mejora cuenta con un servicio de ayuda no presencial llamado “Hora Segura”, el cual comenzó a través del servicio de mensajería de Facebook y tiene por objetivos dar contención, orientación y, en casos puntuales, derivación. En 2015 se logró una subvención de Google para tener un anuncio en el buscador cada vez que alguien escribiera la palabra “formas de suicidarme”, “suicidio sin dolor”, etc. Con esta campaña se incrementó exponencialmente la demanda por el servicio “Hora Segura”, de un promedio de 5 conversaciones semanales a 5 conversaciones diarias en 2016. En 2017 logran una alianza con Let’s Talk para lanzar una aplicación de mensajería, y generar una plataforma que permite la gestión de la atención, sin importar que las solicitudes de conversación lleguen desde la aplicación, desde el chat de Facebook de la Fundación o desde el chat de su página web. Actualmente la atención se brinda en ciertos horarios:

¹ La descripción de este proyecto fue creada por el GobLab de la Universidad Adolfo Ibáñez como parte de un curriculum abierto en formulación de proyectos de ciencia de datos desarrollado en conjunto con la Universidad de Chicago, Carnegie Mellon University e ITAM. Conoce más aquí:

<https://goblab.uai.cl/proyecto-curso-ciencia-de-datos-para-directivos-publicos/>

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.

³ Ministerio de Salud Chile (2019), Programa Nacional de Prevención del Suicidio

⁴ Ministerio de Salud de Chile, Departamento de Estadísticas e Información de Salud





lunes a viernes de 13:00 a 14:00 horas, y de 18:00 a 00:00 horas; domingos de 13:00 a 14:00 horas y de 19:00 a 23:00 horas. Además, hay quienes intentan conectarse con el equipo de voluntarios fuera de esta “Hora Segura”, solicitudes que serán atendidas en el siguiente turno. El equipo está compuesto por profesionales de las ciencias sociales, la salud y la educación, que donan voluntariamente una hora a la semana para dar esta atención. . El problema actual es que el número de usuarios demandando ayuda supera en momentos a los voluntarios disponibles para esta tarea, no pudiendo priorizar contacto con aquellos niños, niñas y adolescentes en alto riesgo de suicidio. Todo Mejora incluso ha identificado factores externos que afectan la demanda por el servicio de ayuda, por ejemplo, la visibilidad del servicio en el marco de la serie “Por Trece Razones” (13 Reasons Why) de Netflix, serie que habla de un caso de bullying y acoso que finalmente termina en suicidio.

Específicamente, Todo Mejora cuenta con un equipo de entre 60 y 80 profesionales, los cuales deben responder entre 8.000 y 9.000 solicitudes de conversación al año. Anualmente, se reciben alrededor de 60% de consultas desde adolescentes de la comunidad LGBTIQ+, y 40% de adolescentes heterosexuales cisgénero. 68% de las personas que acuden a la “Hora Segura” presenta comportamiento suicida en los últimos dos meses.

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

Niños, niñas y adolescentes (NNA) que buscan contención y orientación. Familia y entorno cercano de los mismos NNA.

3.3 ¿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.)

Tomando como NNA al grupo etario entre 12 y 18 años, en Chile hay 1.677.106 niños, niñas y adolescentes según el Censo 2017. Chile, junto con Corea del Sur, son los únicos países donde los suicidios de adolescentes han ido aumentando en vez de disminuir. En el año 2017 se suicidaron 1.878 personas en Chile, de las cuales 93 corresponden a NNA entre 10 y 19 años. Entre los 20 y 24 años este número aumenta a 196.⁵ El 60% de los usuarios de Hora Segura tiene 19 años o menos, 24% tiene entre 20 y 24 años, 9% entre 25 y 29 años y el 8% tiene más de 30 años.

⁵ Revista el Sábado. “El silencio de los suicidios en Chile” (2019)





Entre agosto de 2017 y julio de 2018 se recibieron 8.714 consultas, de las cuales 6.902 fueron concretadas, y dentro de las que se pueden identificar a 2.428 usuarios distintos.

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

Porque la misión de Todo Mejora es promover el bienestar de niños, niñas y adolescentes que sufren de bullying y comportamiento suicida. El mayor foco está puesto en NNA que sufren de discriminación basada en su orientación sexual, identidad y/o expresión de género.

4. 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 los casos de alto riesgo que no son atendidos por falta de tiempo de voluntarios.	Existe una cantidad limitada de voluntarios de atención
2	Disminuir los casos de alto riesgo suicida que no son atendidos por ser consultas que ingresan fuera de horario de atención.	El programa sólo tiene 40 horas de atención a la semana.
3		





5. 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>	Establecer comunicación con el/la NNA que acceden a la plataforma en “Hora Segura”	Establecer comunicación con el/la NNA que requiere de ayuda fuera de la “Hora Segura”	
¿Quién ejecuta la acción? <i>P. ej., departamento de TI (envía correo electrónico)</i>	Equipo de voluntarios	Equipo de voluntarios	
¿Sobre quién/qué se está realizando la acción? <i>P. ej., propietario de automóvil</i>	Usuarios de la app, del chat de Facebook o del chat de la web de Todo Mejora, que hablan dentro de los horarios establecidos	Usuarios de la app, del chat de Facebook o del chat de la web de Todo Mejora, que hablan fuera de los horarios establecidos	
¿Con qué frecuencia se toma la decisión de realizar esta acción? <i>P. ej., anualmente</i>	<i>En los horarios establecidos de “Hora Segura”</i>	Cuando se inicia un nuevo turno después de un período de interrupción.	
¿Qué canales se están usando/se pueden usar para realizar esta acción?	<i>App “Hora Segura”, Facebook chat, chat web</i>	<i>App “Hora Segura”, Facebook chat, chat web</i>	





P. ej., correo electrónico			
Otra información útil sobre la acción		No siempre los voluntarios tienen tiempo para atender inmediatamente las solicitudes de apoyo cuando comienza un nuevo turno, y no siempre es posible reestablecer el contacto.	

6. 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)

	Registro de mensajes entre NNA y sicólogos a través de la App	Etiquetas de segmentos de conversaciones	Fuente de datos 3
¿Qué contiene? P.ej., registros de admisión y alta de hospitales a nivel nacional	Mensajes históricos entre el usuario y el/la voluntario/a.	Clasificación de riesgo de segmentos de conversación	
¿Qué nivel de granularidad? P. ej., transacción, persona, organización, ubicación	Mensajes recibidos y enviados por los usuarios cada vez que hacen uso de la plataforma	Segmento de conversación	





<p>¿Con qué frecuencia se recopila/actualiza una vez que se capta? P. ej., en tiempo real, a diario, semanalmente, mensualmente, anualmente, de manera excepcional</p>	En tiempo real	Una vez, para la construcción del algoritmo.	
<p>¿Tiene identificadores únicos y confiables que se pueden vincular con otras fuentes de datos? P. ej., RUN</p>	No, solo se cuenta con correo electrónico de quienes se contactan por la App.	No aplica	
<p>¿Quién es el propietario interno de los datos? P. ej., hospitales</p>	Todo Mejora	Todo Mejora	
<p>¿Cómo se almacena? P. ej., en una base de datos, PDF, Excel</p>	Base de datos	Base de datos	
<p>Comentarios adicionales</p>	Los correos electrónicos pueden ser no verídicos.	Se requiere que un grupo de voluntarios entrenados clasifique el nivel de riesgo de los segmentos de conversación.	

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

<p>¿Qué contiene?</p>			





¿Qué nivel de granularidad?			
¿Con qué frecuencia se recopila/actualiza una vez que se capta?			
¿Tiene identificadores únicos y confiables que se pueden vincular con otras fuentes de datos?			
¿Quién es el propietario interno de los datos?			
¿Cómo se almacena?			
Comentarios adicionales			

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.)

Sería interesante poder contar con información del ministerio de salud sobre las atenciones médicas y psicológicas de los niños, niñas y adolescentes para complementar la base de datos de atenciones. Sin embargo existe una limitación, no muchos NNAs buscan ayuda tradicional por miedo a más rechazo.

Adicionalmente, si se pudiera hacer un cruce entre los comentarios en Facebook/twitter, fotos de Instagram para encontrar otros factores de riesgo.





7. 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)	Detección		
Propósito del análisis (p. ej., entender el comportamiento histórico de las personas, estimar el riesgo de enfermedad de un paciente)	Detectar riesgo de suicidio de un NNA que accede a la comunicación con voluntarios a través de la App.		
¿Qué acción respaldará este análisis?	Acciones 1 y 2		
¿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. 1.794 segmentos (pares) de conversación, de los cuales 612 fueron evaluados de alto riesgo de suicidio. Se usarán 1.079 segmentos como set de entrenamiento, y los		





	sets de validación y testeo serán de 358 segmentos cada uno.		
--	--	--	--

8. Consideraciones éticas

<p>Privacidad ¿Estás trabajando con datos personales y/o confidenciales identificables a nivel individual?</p>	<p>Sí, a través del correo electrónico se puede deducir la identidad de una persona. El contenido de las conversaciones puede revelar datos personales (ej domicilio) y datos sensibles (salud mental y otros).</p>
<p>Transparencia ¿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>Directora Social de Todo Mejora del uso de los datos en el algoritmo y la persona que coordina a los voluntarios/as del programa Hora Segura para evaluar la aplicabilidad del algoritmo.</p>
<p>Discriminación/equidad ¿Existen grupos específicos para los que deseas garantizar la equidad de los resultados?</p>	<p>Sí, los niños y jóvenes LGTBI+ tienen un mayor riesgo que otros y son un grupo prioritario para la fundación.</p>
<p>Licencia social Si la población completa del país se entera de tu proyecto, ¿lo aprobará?</p>	<p>Puede ser complejo ya que estaríamos “priorizando” la atención en base a un algoritmo. Hay que tener mucho cuidado con los falsos negativos, quienes pueden quedar sin atención. Se podría agregar un segundo indicador en base a tiempo de espera, para asegurar atención a todos.</p>
<p>Rendición de cuentas ¿Quiénes son las personas responsables de todos los elementos anteriores?</p>	<p>Todo Mejora al ser propietario de los datos (conversaciones)</p>





Otras consideraciones, como consentimiento, legales, etc.	
--	--

9. ¿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.

Una vez diseñado el algoritmo de predicción, se pueden realizar distintos ensayos para medir la eficacia del instrumento: por ejemplo, preguntar a los voluntarios si encontraban que el nivel de riesgo detectado por el algoritmo va en concordancia con el nivel de riesgo identificado por ellos.

10. ¿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
Todo Mejora	Supervisará las etapas del proyecto	Directora Social, Rocío Faúndez
Todo Mejora	Implementación del algoritmo	Coordinación de equipos de voluntarios de “Hora Segura”
Let’s Talk	Elaboración del algoritmo predictivo y validación	Equipo de análisis de datos

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.

