

## **Project Scoping**

**1. Nombre del Proyecto:** Llamadas de Emergencia

**2. Nombre de la Organización:** Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la Ciudad de México (C5 para abreviar).

**3. Descripción del Problema:**

### **3.1. ¿Cuál es el problema que están enfrentando?**

Existen más llamadas reportando incidentes viales que ambulancias disponibles. El problema está en que es necesario determinar cuáles llamadas son realmente urgentes con el fin de asignar eficientemente los recursos y principalmente evitar usar una ambulancia en una llamada falsa (por ejemplo, una broma telefónica).

### **3.2. ¿Quién o qué se ve afectado por este problema?**

Las personas que más se ven afectadas son, evidentemente, aquellas involucradas en un accidente vial que no pueden ser atendidas por falta de ambulancias disponibles. Sin embargo, las repercusiones van más allá. Los paramédicos y doctores involucrados en el patrullaje de las ambulancias se ven negativamente afectados por las llamadas falsas dado que podrían haber usado ese tiempo de mejor forma. Finalmente, la sociedad en su conjunto se ve perjudicada a largo plazo porque si se expande la creencia de que no va a llegar una ambulancia a tiempo si es que incurres en un accidente, puede que la gente deje de llamar y que se pierdan vidas en el proceso.

**3.3. ¿Cuánta gente/organizaciones/lugares/etc se ven afectados y qué tanto se ven afectados?**

Con base en los datos proporcionados, sabemos que aproximadamente 20.41% de las llamadas totales recibidas por el C5 son etiquetadas como “falsas” o “negativas”, definiendo llamadas falsas como que, al momento de llegar al lugar de los hechos, el incidente se considera falso (fue una llamada errónea, como una broma); las llamadas negativas son aquellas en donde no se confirma la emergencia o la solicitud de apoyo.

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

Porque el C5 está consciente de la importancia que tiene la asignación eficiente de los recursos escasos que están a su cargo y el potencial que tiene esta asignación en afectar —o hasta determinar— la vida de las personas involucradas en accidentes viales en la CDMX .

**4. Metas. ¿Cuáles son las metas de política pública que serán logradas al resolver este problema y cuáles son las restricciones presentes? (en orden de prioridad)**

	<b>Meta</b>	<b>Restricciones</b>
<b>1</b>	Maximizar la precisión con la que se detectan llamadas que realmente requieren una visita de ambulancia. (donde una predicción positiva implica que sí hay necesidad de una ambulancia y una negativa lo contrario)	No es posible comprobar la exactitud de todas nuestras predicciones dado que solo tenemos veinte ambulancias para enviar en caso de un incidente y pueden reportarse muchas más llamadas.
<b>2</b>	Garantizar el uso eficiente de los servicios de emergencia, para la adecuada atención médica de los incidentes viales.	Misma restricción que en el inciso anterior.
<b>3</b>	Mejorar la experiencia de trabajo de los paramédicos, doctores y demás operadores de ambulancias al reducir el número de llamadas falsas o negativas atendidas.	Misma restricción que en el inciso anterior.

## 5. Acciones

	<b>Acción 1</b>	<b>Acción 2</b>
<b>Acción:</b>	Predecir si una llamada es falsa o no.	Ir al lugar donde se reportó un accidente vial
<b>¿Quién va a ejecutar la acción?</b>	El sistema que diseñen los científicos de datos (que en teoría lo podrán usar todos los que lo requieran)	Las personas que trabajan en el patrullaje (paramédicos, conductores, etc.)
<b>¿Sobre quién o qué se ejecuta la acción?</b>	Sobre las características particulares de la llamada y el incidente reportado (hora del día, día de la semana, delegación, etc.)	Sobre las personas involucradas en el accidente, si es que lo hubo.
<b>¿Qué tan seguido se toma la decisión de realizar esta acción?</b>	Cada que se recibe una llamada acerca de un accidente vial.	Cada que se recibe una llamada acerca de un accidente vial.
<b>¿Qué</b>	Medios digitales.	En persona.

<b>canales se utilizan para tomar esta acción?</b>		
<b>Información adicional relevante</b>	El modelo que diseñen los científicos de datos deberá poder emitir una predicción en tiempo real.	Hay distintas formas de contactar al centro de emergencias (teléfono, botón de auxilio, aplicativo, etc.)

## 6. Los datos

### A. ¿Cuáles fuentes internas de datos tienes?

	<b>Fuente de datos 1</b>	<b>Fuente de datos 2</b>	<b>Fuente de datos 3</b>
<b>Nombre</b>	Base de datos de accidentes viales en la CDMX		
<b>¿Qué contiene?</b>	Información general acerca de cada incidente vial reportado en los últimos años. Incluye fecha de creación y de cierre del reporte (a nivel de hora, día, mes y año), latitud y longitud del incidente, clasificación del incidente, alcaldía, etc.		
<b>¿Cuál es el nivel de granularidad?</b>	La base de datos está a nivel incidente. Hay un registro por cada llamada atendida.		
<b>¿Qué tan frecuentemente se captura o se actualiza?</b>	La información se actualiza diariamente.		
<b>¿Tiene identificadores</b>	No. El identificador único del folio		

<b>únicos que pueden ser conectados con otras fuentes de datos?</b>	únicamente sirve dentro del contexto de este problema.		
<b>¿Quién es el dueño interno de los datos?</b>	El C5		
<b>¿Cómo está almacenada la información?</b>	En un archivo csv.		
<b>Comentarios adicionales</b>	Actualmente se cuentan con más de un millón de registros por analizar.		

#### **B. ¿Cuáles datos pueden ser obtenidos de fuentes externas (públicas o privadas?)**

Para este proyecto, se trabajará sólo con la información disponible en la base de datos de accidentes viales de la CDMX. Sin embargo, existe una fuente adicional que vale la pena considerarse.

	<b>Fuente 1</b>	<b>Fuente 2</b>	<b>Fuente 3</b>
<b>Nombre</b>	Base de datos del sistema de video de la CDMX.		
<b>¿Qué es lo que contiene?</b>	Grabaciones en tiempo real de distintas locaciones en la CDMX.		
<b>¿Cuál es el nivel de granularidad?</b>	A nivel calle (probablemente solo una fracción de la calle)		
<b>¿Con qué frecuencia es recolectado/actualizado?</b>	En tiempo real.		
<b>Tiene identificadores</b>	Desconocido.		

<b>confiables y únicos que pueden ser conectados a otras fuentes de datos?</b>			
<b>¿Quién es el propietario interno de los datos?</b>	El Gobierno de la CDMX.		
<b>¿Cómo se almacena?</b>	-		
<b>Comentarios adicionales</b>	Esta base de datos podría ayudar a confirmar que en efecto no hay falsos negativos en las predicciones.		

**C. En un mundo ideal, ¿qué datos adicionales te gustaría tener/recolectar que serían relevantes para el problema?**

Sería de gran utilidad contar con bases de datos similares de otras ciudades o de organizaciones de servicio médico privado, que nos ayudaran a entender la proporción de llamadas falsas y verdaderas, con el mismo tipo de variables explicativas que ya tenemos.

Sería útil también conocer modelos realizados en otros países o instituciones que intentan analizar fenómenos similares.

## 7. Análisis

	<b>Análisis 1:</b>	<b>Análisis 2:</b>	<b>Análisis 3:</b>
<b>Tipo de Análisis</b>	Clasificación.		
<b>Propósito del análisis</b>	Se asignará la etiqueta "1" cuando la llamada es falsa o negativa y "0" cuando no lo es.		

<b>¿Qué acción será informada por este análisis?</b>	El envío de una ambulancia al lugar de los hechos. Si la llamada se clasifica como falsa o negativa, no se enviará una ambulancia.		
<b>¿Cómo se validará el análisis utilizando datos existentes?</b>	<p>Empleando los datos históricos de llamadas recibidas y atendidas por el C5, tomando en cuenta las llamadas clasificadas como falsas. Hay varias formas de evaluar la utilidad del modelo:</p> <ul style="list-style-type: none"> <li>- Comparar con un modelo nulo donde todas se predicen como llamadas donde sí vale la pena ir.</li> <li>- Comparar con un modelo aleatorio donde con una cierta probabilidad se predice si debe ir o no una ambulancia.</li> <li>- Evaluar con el método tradicional de entrenamiento/prueba o validación cruzada.</li> </ul>		
<b>¿Qué metodología y métricas se utilizarán?</b>	<p>Mediremos la precisión de nuestro modelo de clasificación. Utilizaremos validación cruzada.</p>		

## 8. Consideraciones éticas

<p><b>Privacidad</b></p> <p>¿Se trabaja con datos personales y/o sensibles que pueden ser identificados identificables? Mencionarlos</p>	<p>No.</p>
<p><b>Transparencia</b></p> <p>¿Qué <i>stakeholders</i> deben estar informados sobre qué partes del proyecto?</p>	<p>Capturistas de datos de las llamadas de emergencia, para garantizar que la información registrada sea veraz, y por ende tener una predicción precisa.</p> <p>Autoridades del C5, Protección Civil, policiacas y de servicios de emergencia locales, sobre el hecho de que habrá llamadas que se clasificarán como falsas de entrada.</p> <p>Departamento de T.I. para el entendimiento del producto de datos desarrollado.</p> <p>Los paramédicos y demás personal involucrado en las ambulancias deben tener una idea de cómo funciona el modelo y qué busca lograr.</p> <p>Las personas que hacen las llamadas de emergencia (en caso de que se agregue un anuncio preliminar en las llamadas del estilo de “esta llamada puede ser monitoreada con fines de calidad”)</p>
<p><b>Discriminación/Equidad</b></p> <p>¿Existen grupos específicos para quienes se busca asegurar equidad en los resultados?</p>	<p>Existe un riesgo de descuidar calles, carreteras, caminos, delegaciones o zonas difíciles de acceder (como un barranco) o con poca frecuencia en la base de datos.</p>
<p><b>Licencia Social</b></p> <p>Si toda la población del país se enterara del proyecto, ¿estarían de acuerdo con él? ¿Por qué?</p>	<p>Sí, porque la mayoría de la población estaría de acuerdo en que es un desperdicio de recursos enviar ambulancias a zonas donde se requieren, sobre todo cuando solo hay 20 disponibles.</p>
<p><b>Responsabilidad</b></p> <p>¿Quiénes son las personas responsables por todo lo establecido anteriormente?</p>	<p>C5 de la CDMX. Capturistas de datos de las llamadas de emergencia. Consultores de ciencia de datos.</p>
<p><b>Otras consideraciones como consentimiento, leyes, etc.</b></p>	<p>El modelo debe contar con la precisión suficiente para seguir garantizando la atención médica en accidentes viales de la población y ganarse la confianza en el C5 de que es una mejor alternativa que los modelos nulos o las soluciones actuales.</p>

**9. ¿Qué prueba de campo o prueba aleatorizada controlada puedes diseñar para validar el proyecto en el campo?**

En primer lugar se deberá continuar con el registro de todos los incidentes atendidos por ambulancias y su clasificación por medio del código de cierre, para hacer referencia con lo reportado por el modelo de clasificación.

Adicionalmente, se podría despachar de manera aleatoria ambulancias a llamadas clasificadas como falsas, para garantizar que en realidad se trataba de una llamada falsa y no había emergencia que requiriera atención médica.

**10. ¿Quiénes son las organizaciones externas y los departamentos internos que deben estar involucrados?**

<b>Organización/Departamento</b>	<b>Descripción del involucramiento deseado</b>	<b>Nombre / Rol de Contraparte</b>
C5 de la CDMX	En todas las etapas del proyecto.	Tomadores de decisiones.
Protección Civil	Conocimiento de la forma de operar del modelo de clasificación.	Encargado de informar a los servicios de emergencia.
Departamento de T.I.	Implementación de herramientas necesarias.	Facilitador de servicios tecnológicos e infraestructura.