

# Proyecto de Scoping

## Worksheet: Distribución de ambulancias

### Supuestos

1. Contamos con 20 ambulancias disponibles
2. El C5 verifica la veracidad de las llamadas a través de cámaras o realiza una verificación presencial y lo registra en la base de datos

### Scoping

1. Nombre del proyecto: Accidentes C5
2. Nombre de la organización: AmbulanciaSOS
3. Descripción del proyecto:

a. ¿Cuál es el problema a resolver?

Contamos con un número limitado de ambulancias y llamadas de emergencia que requieren una. Muchas de las llamadas son falsas o repetidas, por lo que se desperdician recursos en atención a casos verdaderos cuando se envía una de las ambulancias a un lugar que no lo necesita.

b. ¿Quién/qué está siendo afectado por el problema?

Las personas que requieren de atención médica urgente y que no llega una ambulancia a darles atención, o llega demasiado tarde.

c. ¿Cuántas organizaciones/personas/lugares están siendo afectadas y qué tanto?

- Personas que requieren atención médica de urgencia
- El C5, por desperdicio de recursos en cuanto las horas de las ambulancias y los médicos que son remunerados
- Las personas que van en las ambulancias al arriesgarse por atender de la forma más rápida una urgencia

d. ¿Por qué es una prioridad resolver el problema para la organización?

Al estar hablando de la posibilidad de que la vida de una persona esté en riesgo y que existe una alta demanda de ambulancias para emergencias, es impermissible que se haga un mal uso de los recursos disponibles.

4. Objetivos: ¿cuáles son los objetivos del negocio/política que serán logrados al resolver el problema y qué limitaciones tienen?

Goal	Constraints	Prioridad (lo definimos el domingo)
Atender la mayor cantidad de emergencias verdaderas	-Mal uso de las personas de la línea 911 -Falta de cobertura total en cámaras o recursos que permitan confirmar los casos a distancia para conocer si los casos son verdaderos	1
Optimizar el uso de recursos del C5	-Presupuesto y ambulancias limitadas -Error al momento de ejecutar la estrategia	2

5. Acciones

	Acción 1 (minimizar violaciones)	Acción 2 (optimizar recursos)	Acción 3
Acción	Sistema de predicción de casos verdaderos para el envío de unidades	Capacitación del call center para obtener más información sobre los incidentes	Campaña de concientización de la sensibilidad de llamadas al 911
¿Quién ejecuta la acción?	C5	C5	Gobierno de la CDMX
¿En quién se tomará la acción?	En los casos determinados como verdaderos	Las personas que llaman a la línea de emergencia 911	La población en general
¿Con qué frecuencia se toma la decisión	Cada vez que el modelo arroje un "caso verdadero"	Cada 6 meses se buscará realizar la capacitación	Una vez al año

de tomar esta acción?			
¿Qué canales se pueden usar para realizar la acción?	<p>Digital: para la notificación del caso con banderas de casos posiblemente falsos</p> <p>Radio: llamadas de confirmación de los casos de las unidades enviadas</p>	Grupos de capacitación	Medios de comunicación masivos
Otra información relevante			

## 6. Datos

### a. Fuentes internas

	Fuente de Datos para la acción 1	Fuente de Datos para la acción 2
Nombre	1. Reportes realizados en el pasado con etiqueta de casos falsos, informativos o duplicados anteriores	1. Base de datos de las personas que labora en el call center
¿Qué contiene?	1. Motivo de la llamada, folio, geolocalización, delegación, día, hora, año y etiqueta de caso verdadero, falso, informativo o duplicado.	1. Nombre completo, número de empleadx, tiempo de trabajo en el call center, jefe directo, horas y días laborales
¿Qué nivel de granularidad tienen los datos?	<p>Tenemos información desagregada, desde nivel de procedencia de la emergencia:</p> <ol style="list-style-type: none"> <li>Geolocalización</li> <li>Día y hora del hecho</li> </ol>	Nivel trabajador(a)
¿Qué tan	Dlaria	Cada mes

frecuente se colectan y/o actualizan después de ser capturados?		
¿Tiene identificadores únicos y seguros que puedan ser ligados a otras fuentes de datos?	La información cuenta con la granularidad mínima para realizar cruces entre diferentes fuentes de información mediante: 1. - ID folio	Número de empleadx
¿Quién es el dueño interno de los datos?	C5	C5
¿Cómo se almacena?	Base de datos consumible por servicio web	Base de datos renovable cada mes
Comentarios Adicionales.		

b. Fuentes externas, privadas y/o públicas

	Fuente de Datos para la acción 1
Nombre	Shapefiles CDMX
¿Qué contiene?	1. Datos de la división geográfica de la Cuidad de México
¿Cuál es el nivel de granularidad?	Nivel colonias
¿Qué tan frecuente se colectan y/o actualizan después de ser capturados?	Cada vez que el Instituto Electoral de la CDMX modifique la base de datos

¿Tiene identificador es únicos y seguros que puedan ser ligados a otras fuentes de datos?	1. ID colonia
¿Quién es el dueño interno de los datos?	Instituto Electoral de la CDMX
¿Cómo se almacena?	1. Shapefile
Comentarios adicionales	

c. En un mundo ideal, ¿hay datos adicionales que te gustaría tener/conseguir que sería relevante para el problema?

- Base de datos con información de la localización de donde salen las ambulancias
- Registro de las unidades médicas más cercanas al lugar de la emergencia
- Base de datos con información de cómo se identifica que una emergencia es verdadera (cámaras, presencial, radio, etc...)
- Registro de placas de las ambulancias de SEMOVI (Secretaría de Movilidad)

## 7. Análisis

	Análisis 1	Análisis 2	Análisis 3
Tipo de Análisis	Clasificación  Que las ambulancias que se lancen a un caso sean más propensos a atender casos verdaderos	Random forest para seleccionar las variables importantes	Impacto de la campaña de concientización

Propósito del análisis	Analizar los patrones que pueden ayudar a identificar llamadas falsas	Eliminar variables que no generan información relevante para la identificación de casos verdaderos	Identificar si la campaña en medios desincentiva a la población a realizar llamadas de broma al 911
¿A cuál acción informa este análisis?	Acción 1	Acción 2	Acción 3
¿Cómo validarías el análisis utilizando los datos existentes? ¿Cuál metodología y cuáles métricas utilizarías?	Tasa de casos atendidos que sean verdaderos con backtesting  Comparamos con el baseline de tasa de detección de anomalías actual	Error del modelo	Porcentaje de llamadas de broma después de haber identificado casos falsos

## 8. Consideraciones Éticas

Privacidad	<ul style="list-style-type: none"> <li>- Datos personales de las personas que piden la atención (teléfono)</li> <li>- Datos personales de las personas que trabajan en el call center</li> </ul>
Transparencia	<ul style="list-style-type: none"> <li>- Fase 1: Construcción y definición de modelo</li> <li>- Fase 2: Test del modelo (DS+Testers)</li> <li>- Fase 3: Estrategia final (Stakeholders del C5)</li> <li>- Sólo tendrán acceso a la información los científicos de datos (en todas las partes del análisis) involucrados en el proyecto, el C5 y los tomadores de decisiones de la entidad gubernamental (resultados del modelo)</li> </ul>
Discriminación/Equidad	<ul style="list-style-type: none"> <li>- Falsos negativos, se debe de tener en cuenta que es posible que el modelo etiquete como falsa una llamada verdadera, pero el modelo permite hacer una minimización del daño a</li> </ul>

	través de hacer más eficientes los recursos disponibles.
Licencia Social	<ul style="list-style-type: none"> <li>- Riesgo de que los stakeholders no quieran asumir el riesgo de no atender un falso negativo.</li> <li>- Riesgo reputacional para el 911 de no atender un caso grave por atender a los resultados del modelo.</li> </ul>
Responsabilidad	<ul style="list-style-type: none"> <li>- Generadores de estrategia, ejecutores y equipo de liderazgo, Gobierno de la CDMX y C5.</li> </ul>
Otras consideraciones	

9. ¿Qué ensayo de campo o ensayo controlado aleatorio puede diseñar para validar el proyecto en el campo? Los resultados que medirá deben coincidir con sus objetivos. Defina la población en la que se probará el modelo. Defina la duración de la prueba. Especifique la línea de base. Debe medir el impacto en diferentes subgrupos de población.

Se deberá de probar el modelo con un test en tiempo real cuando esté terminado. La propuesta es que durante un mes esté corriendo la detección de falsos del modelo mientras el C5 continúa con su operación actual, de esta forma se identificará la eficacia del modelo comprobando las etiquetas que este arroja contra las detecciones del mecanismo actual. Para la validación en campo tenemos las cámaras de C5 y las verificaciones personales que se realizan a la hora de recibir una llamada solicitando una ambulancia. El baseline será la tasa de detecciones falsas con el mecanismo actual contra las detectadas con el modelo. Los subgrupos de inicio serían las alcaldías.

10. ¿Quiénes son las organizaciones externas y departamentos internos que necesitan estar involucrados?

Organización/Departamento	Descripción del involucramiento	Nombre/rol de contraparte
Sistema de	Proporcionar	Asignador de

ambulancias de la CDMX	infraestructura médica Aportar información sobre los casos atendidos	ambulancias
Servicios de salud de la CDMX	Atención a las ambulancias enviadas en urgencias	Directorxs de urgencias de las unidades de salud
Departamento de TI de C5	Proveer infraestructura para el almacenamiento y la manipulación de datos	Directorx de TI del C5
Departamento de estadística	Proveer datos sobre la población atendida	Directorx de estadística del C5