

Data Science Project Scoping

1. Nombre del proyecto:

Clasificación de reportes falsos de Incidentes Viales reportados en el C5.

2. Nombre de la Organización:

El Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la CDMX (C5).

3. Descripción del problema:

“El 23 de diciembre de 2015 el Gobierno de la Ciudad de México decretó la anexión de LOCATEL al Centro de Atención a Emergencias y Protección Ciudadana de la Ciudad de México (CAEPCCM) para conformar el Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la Ciudad de México “C5”.

El C5 es la dependencia del Gobierno de la Ciudad de México encargada de captar información integral para la toma de decisiones en materia de seguridad pública, urgencias médicas, medio ambiente, protección civil, movilidad y servicios a la comunidad en la capital del país a través del video monitoreo, de la captación de llamadas telefónicas y de aplicaciones informáticas de inteligencia.

Actualmente, los siguientes servicios del C5 operan las 24 horas, los 365 días del año:

- **Video Monitoreo:** el C5 opera y monitorea las más de 15 mil cámaras de vigilancia que hay en la Ciudad de México (Sistema Tecnológico de Videovigilancia STVs), con la finalidad de prevenir y alertar inmediatamente a las autoridades de seguridad y de emergencias capitalinas sobre cualquier situación de riesgo.
- **Servicio de Atención de llamadas de emergencia 9-1-1 CDMX:** se atienden y canalizan de manera inmediata los reportes de delitos y emergencias como incendios, robos, accidentes automovilísticos, urgencias médicas, entre otros.
Cuando se da a conocer alguna emergencia por la aplicación (App 9-1-1), los despachadores del C5 tienen la posibilidad de saber la ubicación del reporte, lo que facilita la precisión en el envío de los cuerpos de emergencia. Cuenta con la opción de enviar imágenes de la emergencia, las cuales llegan de manera directa al despacho del C5. Así mismo las y los usuarios pueden crear una red de emergencia con familiares y conocidos, que al ser activada, serán alertados en caso de un evento extraordinario.
- **Denuncia Anónima 089:** se atiende a los ciudadanos que desean reportar, de manera anónima, algún acto ilícito que se ha cometido en su entorno como narcomenudeo, secuestro, violencia intrafamiliar, entre otros. Los reportes de las y los ciudadanos son canalizados inmediatamente a las autoridades correspondientes las cuales inician la investigación y en su caso la persecución del delito.
- **LOCATEL 5658 1111.**

A partir de estos canales de entrada de reportes, se ha detectado que para los incidentes viales existe un porcentaje elevado de reportes que son considerados como falsos, pues las unidades de emergencia llegan al lugar del reporte y no encuentran indicios de incidentes. Particularmente, el problema al que nos enfrentamos es que el C5 sólo cuenta con 20 ambulancias para enviar en caso de un incidente. Dados los recursos limitados, la problemática radica en el envío de ambulancias a llamadas/reportes que resultan falsos.

3.1. ¿Cuál es el problema que estamos enfrentando?

Nos enfrentamos dos problemas: el primero son los recursos limitados pues sólo se cuenta con 20 ambulancias para enviar en caso de un incidente; y el segundo es que existe una cantidad considerable de reportes falsos (20% del total de incidentes reportados). Lo anterior provoca la posibilidad de pérdida de vidas humanas que realmente necesitan de una ambulancia pero que se les fue impedida por desviar los escasos recursos a reportes falsos.

Por lo tanto, se tiene que priorizar el envío de recursos a llamadas que sean determinadas como verdaderas. Es decir, es necesario identificar los reportes falsos con la mayor precisión para no enviar ambulancias a esos incidentes falsos.

3.2. ¿Quién o qué se ve afectado?

1. La ciudadanía que necesita una ambulancia, pero no la recibe dado que ésta atendió a una llamada falsa.
2. El gobierno de la Ciudad de México y el C5 , quienes destinan recursos a las llamadas que son falsas.

3.3. ¿Cuántas personas/organizaciones/lugares/etc. y qué tanto se ven afectados?

1. Organizaciones:

1.1. El gobierno de la CDMX y el C5: entidades que mandan las ambulancias, principalmente del Escuadrón de Rescate y Urgencias Médicas (ERUM). El ERUM da atención médica prehospitalaria, ayuda y rescate en casos de accidentes, auxilio a mujeres en labores de parto, apoyo en vías públicas, atención dentro y fuera de eventos masivos, atención a personas accidentadas, traslado de enfermos y lesionados a hospitales. Diariamente se atiende un promedio de 322 emergencias de cualquier caso. De enero a agosto del 2019, el ERUM apoyó a 78,375 personas entre las que destacan lesionados en vía pública, mujeres en labor de parto, menores abandonados, motociclistas accidentados y ciclistas lesionados.¹

2. Personas:

2.1. La ciudadanía de la CDMX. De acuerdo con la información del INEGI sobre el número de accidentes de tránsito terrestre ocurridos en carreteras no federales y registrados por

¹ Secretaría de Seguridad Ciudadana, Escuadrón de Rescate y Urgencias Médicas. Recuperado de <https://www.ssc.cdmx.gob.mx/micrositios/escuadron-de-rescate-y-urgencias-medicas>

los gobiernos municipales, así como el total de víctimas muertas en el lugar del accidente y heridas con lesiones leves y/o graves, en 2019, en la CDMX hubo un total de 10,673 accidentes de tránsito terrestre, un total de 226 muertos y 2,868 heridos.²

De acuerdo con los datos de incidentes viales reportados por la base de datos proporcionada por el C5, se han reportado hasta 676 incidentes viales diarios (año 2018). El número promedio de accidentes por día de 2013 a 2020 se comportó de la siguiente forma:

Año	Promedio del número de accidentes por día
2014	469
2015	506
2016	522
2017	621
2018	676
2019	629
2020	438

Sin embargo, el número de reportes falsos³ por día en un año ha sido hasta del 22.3%, en el año 2016.

Año	Número de reportes falsos por día	Porcentaje de reportes falsos por día
2014	81	17.0%
2015	104	20.4%
2016	118	22.3%

² INEGI Estadística de accidentes de tránsito terrestre en zonas urbanas y suburbanas. Recuperado de <https://www.inegi.org.mx/temas/accidentes/#Tabulados>

³ Se consideran reportes falsos las observaciones del código de cierre como:

- (N) La unidad de atención a emergencias fue despachada, llegó al lugar de los hechos, pero en el sitio del evento nadie solicitó el apoyo de la unidad.
- (F) El operador/a o despachador/a identifica, antes de dar respuesta a la emergencia, que ésta es falsa. O al ser despachada una unidad de atención a emergencias en el lugar de los hechos se percatan que el incidente no corresponde al reportado inicialmente

2017	134	21.5%
2018	145	21.4%
2019	120	19.0%
2020	86	19.3%

Adicionalmente, en el año 2016 un 16.7% de reportes verdaderos requirieron una ambulancia y posiblemente no se les envió⁴, dado que la categoría del incidente indica lesione. En la siguiente tabla se muestra esta problemática para todos los años.

Año	Número promedio de reportes verdaderos, por día, que requerían una ambulancia y no se les envió.	Porcentaje de reportes verdaderos, por día, que requerían una ambulancia y no se les envió.
2014	52	10.8%
2015	74	14.4%
2016	89	16.7%
2017	99	15.9%
2018	105	15.4%
2019	82	12.9%
2020	54	12.0%

3.4. ¿Por qué resolver el problema representa una prioridad para la organización?

El problema representa una prioridad para la organización porque únicamente cuenta con 20 ambulancias para la atención de incidentes y además tiene reportes falsos.

⁴ Se consideran llamadas verdaderas y que posiblemente no se les envió ambulancia las entradas que cumplen estas dos condiciones:

- La columna codigo_cierre tiene alguna de estas 3 categorías:
 - (A) La unidad de atención a emergencias fue despachada, llegó al lugar de los hechos y confirmó la emergencia reportada.
 - (D) El incidente reportado se registró en dos o más ocasiones procediendo a mantener un único reporte (afirmativo, informativo, negativo o falso) como el identificador para el incidente.
 - (I) El incidente reportado es afirmativo y se añade información adicional al evento.

Como se muestra en el inciso anterior, existen reportes verdaderos que requerían de una ambulancia y no que no se les envió. El número de llamadas y/o reportes falsos por día representa atención inoportuna de emergencias y el empleo de recursos para verificar que efectivamente la llamada es falsa.

De acuerdo con cifras de 2017 de la Organización Mundial de la Salud (OMS), cada año mueren en el mundo cerca de 1.3 millones de personas en accidentes de tránsito, y entre 20 y 50 millones padecen traumatismos no mortales causantes de discapacidad. Los accidentes viales, además, constituyen una de las principales causas de mortalidad en todos los grupos etarios, principalmente entre personas de entre 15 y 19 años⁵.

Además, de acuerdo con datos del Instituto Nacional de Salud Pública (INSP), México ocupa el séptimo lugar a nivel mundial y el tercero en la región de Latinoamérica en muertes por siniestros viales, con 22 decesos de jóvenes de entre 15 y 29 años al día, y 24 mil decesos en promedio al año. Los siniestros viales constituyen la primera causa de muerte en jóvenes entre 5 y 29 años de edad y la quinta entre la población general⁶.

4. Objetivos: ¿Cuáles son las metas del negocio que se alcanzarán al resolver el problema y cuáles son sus restricciones? (en orden de prioridad).

#	Objetivo	Restricciones
1.	Clasificar si los reportes de incidentes viales realizados al C5, son falsos o verdaderos.	<ul style="list-style-type: none"> - Recursos limitados en presupuesto y capital humano para verificar que el reporte es falso. - Problemas legales si se clasifica el reporte como falso y no se atiende, cuando realmente era verdadero. - Restricción de recursos escasos del envío de las ambulancias a aquellos reportes determinados como verdaderos. - Credibilidad de la sociedad en el algoritmo/modelo: la sociedad podría exigir la verificación de un humano en el lugar del incidente para todos los reportes ya que no confía en el modelo de clasificación. - Fuerza política: dado que el gobierno de la Ciudad de México

⁵ Organización Mundial de la Salud. 10 datos sobre la seguridad vial en el mundo. [Internet]. Recuperado de: <http://www.who.int/features/factfiles/roadsafety/es/>.

⁶ Treviño S. Consejos para prevenir accidentes viales. Gaceta INSP [Internet]. 2017. Recuperado de: <http://gaceta.insp.mx/?vol=1>.

		jamás a utilizado un modelo para realizar esta tarea específica, existe el riesgo de que no acaten los resultados arrojados por el modelo desarrollado y opten por continuar haciendo las cosas igual que como las hacen ahora.
--	--	---

5. Acciones

	Acción 1
Acción:	En caso de que la llamada sea verdadera, se envía una ambulancia al lugar.
¿Quién ejecuta la acción?	El C5
¿Sobre qué/quién se ejecuta la acción?	El envío de ambulancia al lugar se ejecuta a la llamada determinada como verdadera. Esta acción se ejecuta sobre las posibles víctimas del incidente.
¿Qué tan frecuente se toma la decisión de ejecutar la acción?	Cada que se recibe una llamada verdadera, siempre y cuando haya ambulancias disponibles.
¿Qué canales se utilizan o se pueden utilizar para ejecutar la acción?	Los canales que tienen los ciudadanos para reportar su incidente al C5 son: botón de auxilio, llamada del 066, radio, cámara, redes, llamada del 911, llamada app911, zello, aplicativos. El canal de envío de la ambulancia es el C5. Una vez que se tiene el reporte se verifica la falsedad de la llamada con el envío de un policía.
Otra información útil sobre esta acción.	Solo se puede ejecutar de acuerdo con las 20 ambulancias disponibles.

6. Datos

6.1. ¿Qué bases de datos se tienen internamente?

	Fuente de Datos 1	Fuente de Datos 2
Nombre:	incidentes-viales-c5.csv	Diccionario de Datos de Incidentes Viales.
¿Qué contiene?	Conjunto de datos con los incidentes viales reportados por el Centro de Comando, Control, Cómputo, Comunicaciones y Contacto Ciudadano de la	Descriptor de variables del archivo incidente-viales-c5.csv

	Ciudad de México (C5) desde 2014.	
Nivel de granularidad	Folio de reporte	Variable
Frecuencia de recolección/actualización después de que es capturada.	Mensualmente	Cada que se actualiza o se modifica la base de datos incidentes-viales-c5.csv
¿Tienen identificadores únicos y fidedignos que pueden ligarse a otras fuentes de datos?	<p>La base de datos contiene el folio que es único y fidedigno. Sin embargo, no tenemos información si con el folio es posible asociar la base con otras fuentes de datos.</p> <p>Con la identificación de la alcaldía es posible ligar con otras fuentes de datos externas con información a nivel alcaldía.</p>	<p>Contiene información descriptiva única del archivo incidentes-viales-c5.csv.</p> <p>No tenemos información si es posible ligar este diccionario con otros catálogos o diccionarios de otras fuentes de datos.</p>
¿Quiénes son los dueños internos de los datos?	C5-Gobierno de la CDMX	C5-Gobierno de la CDMX
¿Cómo se almacenan?	<p>De acuerdo con la información pública de los datos abiertos de la Ciudad de México, los datos están bajo licencia de CC BY y están almacenados en los siguientes formatos de texto plano:</p> <ul style="list-style-type: none"> - CSV - JSON - Excel <p>Además, se encuentran en los siguientes formatos de archivo geográfico:</p> <ul style="list-style-type: none"> - Shapefile - KML - GeoJSON <p>Adicionalmente, está disponible una API para poder descargar la información por medio de este servicio.</p>	<p>De acuerdo con la información pública de los datos abiertos de la Ciudad de México, el diccionario de datos está almacenado en un archivo de Excel.</p> <p>No se cuenta con información del almacenamiento interno en el área de TI del C5.</p>

	No se cuenta con información del almacenamiento interno en el área de TI del C5.	
Comentarios adicionales		

6.2. ¿Qué bases de datos se tienen externamente?

En esta etapa, no se utilizarán fuentes de datos externos.

6.3. En un mundo ideal existen datos adicionales que te gustaría recolectar.

1. Datos de WAZE: verificar en tiempo real los accidentes reportados por los usuarios.
2. Datos de las llamadas a las aseguradoras: verificar en tiempo real los accidentes reportados por los asegurados.
3. Datos de la visualización de las cámaras de tiempo real del accidente: verificar en tiempo real las falsas alarmas.
5. Datos de población por delegación para análisis de cobertura de los recursos.

7. Análisis

	Análisis 1
Tipo de Análisis:	Clasificación
Propósito del análisis	Clasificación de llamadas falsas.
¿Qué acción informará este análisis?	Envío de ambulancias para atender incidentes verdaderos.
¿Cómo se validará este análisis utilizando los datos existentes? ¿Qué metodologías y métricas se utilizarán?	Se utilizará la variable “código de cierre” para validar que clasificación que hace el modelo de las llamadas falsas sea correcta. Se aplicarán metodologías de aprendizaje supervisado y con GridSearch se ajustarán hiperparámetros. Se elegirá el modelo con una mejor precision@k para la clasificación de las llamadas/reportes falsos. Con base en nuestra mejor precision@k, se definirá el recall@k (cobertura@k).

8. Consideraciones éticas

	Análisis 1
Privacidad Se está trabajando con datos personales/sensibles que	No tenemos acceso a datos personales en la base de datos de incidentes viales.

pueden ser individualmente identificables? Mencionarlos.	
Transparencia ¿Qué stakeholders deben saber de qué partes del proyecto?	<ol style="list-style-type: none"> 1. El gobierno de la CDMX y el C5 debe de saber todas las partes del proyecto y el funcionamiento del modelo. 2. La ciudadanía debe de tener transparencia de cómo se destinan los recursos, en este caso las ambulancias. Por lo tanto, es importante que la ciudadanía sepa cuáles son las variables para construir el modelo y la forma de decidir cómo clasificar un reporte como falso, de acuerdo con la ingesta de datos.
Discriminación/Equidad ¿Existen grupos específicos para los que se quiera asegurar equidad en los resultados?	<p>Se requiere la equidad en los resultados para todos aquellos reportes verdaderos que necesitan la ambulancia.</p> <p>De acuerdo con los datos, se puede incurrir en discriminación de algunas delegaciones por falta de datos (por ejemplo de Milpa Alta) y por la proporción de llamadas falsas en una misma ubicación (por ejemplo Cuajimalpa).</p>
Licencia Social Si toda la población del país se enterara del proyecto, ¿estarían de acuerdo con él, por qué?	La población estaría de acuerdo con el proyecto porque se está haciendo eficiente el destino de recursos y recurriendo al protocolo del C5 para verificar un reporte en caso de que una llamada sea falsa.
Responsabilidad ¿Quiénes son los responsables de todo lo anterior?	El C5 como coordinador del proyecto.
Otras consideraciones	Dado que nuestra métrica es la precisión y dada la limitante del número de ambulancias, se podría dejar de atender lugares que los incidentes pueden ser verdaderos, si existe falla humana en la verificación del incidente falso.

9. ¿Qué validación de campo o prueba controlada aleatoria puedes diseñar para validar tu prueba piloto?

La prueba controlada o prueba piloto consistirá en que se aplicará la operación regular del C5 para verificar los reportes clasificados como falsos según nuestro modelo. Cuando el modelo clasifique un reporte como falso, se verifica de acuerdo con el protocolo del C5: se revisa el incidente en las cámaras; si no hay cámaras en el lugar del incidente, se envía un policía a verificar el incidente.

10. ¿Cuáles son las organizaciones externas y departamentos internos que deberán involucrarse?

Organización/Departamento	Descripción de la intervención/participación deseada.	Nombre/rol de la contraparte
---------------------------	---	------------------------------

C5	Proporcionará la infraestructura de los datos, información de la operación actual, el modelo nulo y recursos para la implementación del modelo de ciencia de datos.	Dueños de los datos y líder del proyecto.
Data Science Team (Equipo 6 ITAM)	Análisis, construcción e implementación del modelo de ciencia de datos.	Equipo de científicos de datos.