

Predictor de responsabilidad de siniestros viales

22th Agosto 2022

Introducción

Las aseguradoras de automóviles reciben a diario entre 1000 y 5000 descripciones de un siniestro redactadas por el operador del centro de atención al cliente que recibe la llamada, por ejemplo:

“CIRCULANDO POR LA CALLE MANUEL OBARRIO EN DIRECCIÓN OESTE A ESTE, ESTANDO DETENIDO ESPERANDO PARA CRUZAR DEBIDO A QUE SON 5 ESQUINAS, EN ESE MOMENTO UNA MOTO ME EMBISTE DE FRENTE”

En estas, a menudo, se encuentra la información necesaria para determinar si el asegurado en cuestión es el responsable del siniestro o no según la ley de tránsito vigente, actualmente esta comprensión de la descripción se realiza de manera manual por uno o varios operadores que leen e interpretan el corpus asignando una etiqueta según el protocolo interno, dicha tarea conlleva un costo monetario ya que estos lectores no pueden ser utilizados para otras tareas de mayor importancia o prioridad.

En caso de que la responsabilidad del asegurado sea positiva, se disparan una serie de acciones paliativas para resolver el problema antes de que alcance una instancia judicial, si esto no se logra cabe aclarar que la instancia judicial es más costosa, en caso que no haya responsabilidad por parte del asegurado este caso se coloca en una cola de baja prioridad ya que se supone que no avanzará a la instancia judicial.

El desafío

En el dataset adjunto “*siniestros.csv*” se encuentran las descripciones con sus respectivas etiquetas y otros datos que relacionados al siniestro y al asegurado, dichas columnas son:

-
- **Id_evento**: es un código identificador
 - **Id_reclamo_evento**: es un código identificador
 - **Id_reclamo**: es un código identificador
 - **TIPO_RECLAMO**: es un código identificador
 - **FECHA_SINI**: fecha en la que ocurrió el siniestro
 - **LUGAR_OCURRENCIA**: dirección de ocurrencia del siniestro
 - **COD_POST_OCURRENCIA**: código postal del lugar donde ocurre el siniestro
 - **ESTADO_CIVIL**: estado civil del asegurado
 - **FECHA_NAC_ASEG**: fecha de nacimiento del asegurado
 - **SEXO**: sexo del asegurado
 - **OCUPACION_ASEG**: ocupación codificada del asegurado
 - **MODELO_VEHICULO**: año de emisión del vehículo del asegurado
 - **FECHA_PROCESO**: fecha en la que se procesó la designación del asegurado
 - **TIPO_ADMINISTRACION**: determina la responsabilidad inicial del asegurado
 - **DESCRIP_HECHO**: descripción del siniestro en cuestión
 - **HORA_SINI**: hora en la que ocurre el siniestro
 - **LATITUD**
 - **LONGITUD**
 - **TIPO_ADMINISTRACION_FIN**: determina la responsabilidad final del asegurado

Dentro de la columna **TIPO_ADMINISTRACION_FIN (target)** se encuentran 5 valores posibles donde solo la etiqueta **A** implica la responsabilidad de parte del asegurado. Algunas restricciones a tener en cuenta es que *si una misma descripción tiene dos etiquetas diferentes se debe tomar la de mayor gravedad (A)*.

Objetivos

1. Realizar el análisis exploratorio de los datos para determinar qué campos además de la descripción tienen importancia para determinar la responsabilidad del hecho.
2. Realizar un modelo que dada las características seleccionadas determine la responsabilidad del hecho.
3. Realizar un workflow del proceso realizado en los puntos anteriores donde también se reporten los próximos pasos a seguir al momento de disponibilizar dicho modelo al cliente.
4. Disponibilizar en un repositorio todo el desarrollo con un esquema de la estructura deseada del proyecto y la explicación de su abordaje en el “*README.md*” o en un Notebook junto con el resto de los puntos.

Puntos a ser evaluados

Durante la resolución del desafío debe sentirse totalmente libre de utilizar la tecnología que más le guste y Python como lenguaje de programación, pero se tendrá en cuenta:

- Creatividad ante la resolución general y en detalles particulares
- Calidad del código
- Calidad del proceso
- Escalabilidad y automatización (no es necesario hacerlo pero si el planteamiento)
- Fundamento/criterios tenidos en cuenta al momento de tomar decisiones
- Claridad de la explicación

Confidencialidad

Al realizar este desafío usted comprende que todos los datos contenidos en el dataset son confidenciales y queda por tanto sujeto a una obligación legal de confidencialidad, no pudiendo revelar a ningún tercero dato alguno respecto del dataset. En caso de incumplimiento de la presente obligación de confidencialidad, Legal Hub podrá iniciar las acciones legales correspondientes para recuperar el daño sufrido.

Esperamos que disfrute de este desafío tanto como nosotros.

Gracias y buena suerte.