

# Prevenir complicaciones salva vidas

## Proceso de contratación de PEOPL – Caso técnico B:

**Objetivo:** Desarrollar un modelo capaz de predecir qué pacientes con cáncer, interactuando con nuestro chatbot conversacional, tienen un mayor riesgo de sufrir una hospitalización no planificada en los próximos 30 días.

**Datos disponibles:** Nuestro chatbot recopila datos diariamente de los pacientes en forma de texto. Esto incluye información sobre sus **síntomas, su estado emocional, alimentación, su comportamiento, y cambios en medicación o tratamiento**, entre otras cosas que incluso no se han descubierto.

Los únicos datos que se tienen son el ID del paciente, el contenido del mensaje que envió el paciente, la hora a la que envió el mensaje y por otra parte el ID del paciente, el contenido del mensaje que el asistente virtual envió y la hora a la que se envió.

## Propuesta de ejecución:

- **Benchmark y entendimiento del problema:** ¿alguien ya ha hecho modelos parecidos antes? ¿qué resultados obtuvieron? ¿cuáles son las causas de la hospitalización no planificada en pacientes con cáncer? ¿cuál es el impacto? Definir algunas preguntas iniciales nos pueden ayudar a no reinventar la rueda. Además, podrían definir los procesos que debemos seguir para poder alcanzar nuestro objetivo.
- **Análisis de los datos:** A partir de los datos proporcionados (ID del paciente, contenido del mensaje del paciente, hora del mensaje del paciente, ID del paciente, contenido del mensaje del asistente, hora del mensaje del asistente), se espera que usted realice un análisis exploratorio de los datos para **identificar patrones, tendencias y posibles problemas**.
- **Preprocesamiento y limpieza de datos:** Limpiamos y preprocesamos estos datos. Describir y justificar las técnicas utilizadas para limpiar y **preprocesar los datos considerando la naturaleza de los datos (lenguaje natural)**, incluyendo el manejo de los **valores faltantes y los outliers**.
- **Creación de características:** Generamos características basadas en estos datos. Por ejemplo, podríamos resumir la severidad de los síntomas durante los últimos 7 días, el nivel de adherencia al tratamiento de los pacientes, el número de veces que el paciente ha mencionado sentirse deprimido, o el número de noches que han reportado tener problemas para dormir.
- **Construcción de los perfiles de los pacientes:** a partir de las conversaciones que tiene Pipo con los pacientes necesitamos clasificar e identificar los data points que nos permitirán construir los segmentos de pacientes. Estos segmentos van a estar limitados a la información que brindan los pacientes al asistente virtual a través de mensajes de texto. Aquí podremos encontrar datos como: **síntomas, tratamientos, sexo, edad, emociones, intensidad de síntomas, co-morbilidades**, entre otras variables que podrían alimentar nuestro modelo. ¿Qué variables se consideraron más relevantes y por qué?
- **Modelado:** Desarrolle un modelo que nos permita predecir qué pacientes con cáncer tienen mayor riesgo de hospitalización no planificada en los próximos 30 días. Desarrollar un modelo predictivo, explicar la elección del algoritmo y los parámetros utilizados. Describir en detalle cómo funciona el modelo y cómo se ajusta a los datos disponibles.
- **Evaluación del modelo:** define y justifica la forma de evaluación del modelo, ¿cómo sabes que tu modelo es efectivo para los objetivos de negocio? ¿cuál fue tu criterio para escoger esa forma de evaluación?
- **Return of Investment:** ¿de qué forma este modelo podría impactar en la visión Peopl? ¿**de qué forma este modelo impactaría en dar un mejor atención oncológica a nuestros pacientes**?
- **Siguientes pasos y oportunidades de mejora:** ¿cuál sería tu recomendación sobre los siguientes pasos a tomar con este modelo? ¿consideras que podría ser una prioridad para Peopl? ¿recomendarías seguir invirtiendo recursos o detendrás el proyecto en este

punto?¿qué se puede mejorar?

**Entregables:**

- **Comunicación de resultados:** Grábese en un video de máximo 15 minutos explicando el desarrollo del proyecto y sus decisiones para llegar al objetivo. Enviar el video y la presentación en pdf.
- **Repositorio de GitHub:** Todo el código y la documentación se entregan en un repositorio de GitHub que contiene un README detallado que describe el contenido del repositorio, cómo ejecutar el código y los resultados obtenidos.
- **Deadline:** enviar los entregables a [piero@peopl.health](mailto:piero@peopl.health) hasta el miércoles 12 de julio a las 11:59 pm.

Nota: Debido a la naturaleza sensible de los datos, cualquier información que pueda identificar a los pacientes debe ser manejada con la máxima confidencia.