

Caso Técnico PEOPL

Proceso de contratación de PEOPL – Caso técnico A:

Contexto: Peopl busca a mediano plazo tener una intervención proactiva con los pacientes. Es decir, poder anticiparnos a las complicaciones que un paciente pueda tener y reforzar el acompañamiento que ofrecemos. Si bien revisar cada chat que se tiene con el usuario es una tarea poco eficiente, debe haber una manera de poder detectar patrones y una población que tenga más chance de sufrir alguna complicación durante su tratamiento oncológico.

Datos disponibles: Nuestro chatbot recopila datos diariamente de los pacientes en forma de texto. Te proporcionaremos cuatro tablas. La primera contiene los mensajes que las personas envían al chatbot. La segunda contiene los mensajes enviados por el chatbot. La tercera almacena los síntomas reportados por el usuario con un nivel de intensidad del 1-10 (10 el nivel de dolor más alto). La cuarta tabla guarda información diferente información del usuario.

El detalle de los datos por tabla es:

User Messages - Mensajes del Usuario

1. ID Message - id
2. ID paciente - patient
3. Contenido del mensaje - Body
4. hora de envío de mensaje (horario servidor) - creation_date
5. hora de envío mensaje (horario local de usuario) - local_time

Pipo Messages - Mensajes del asistente

1. ID message - id
2. ID paciente - patient
3. Contenido del mensaje - Body
4. Hora de envío de mensaje - creation_date

Symptom

1. id sintoma - id
2. ID paciente - user
3. sintoma reportado - description
4. intensidad del sintoma - intensity
5. fecha de reporte de sintoma - creation_date

User - Info del usuario

1. ID del usuario - phone_number (unir con patient o user de las otras tablas)
2. Creación del usuario en la base de datos - creation_date
3. Tipo de cancer - cancer_type

Objetivo: Realizar un análisis exploratorio de la data presentada con el fin de obtener insights. Si es posible, armar una variable objetivo (target) a predecir.

hint: Considerar los síntomas con mayor frecuencia.

Propuesta de ejecución (puedes proponer una metodología a tu gusto):

- **Análisis de los datos:** A partir de los datos proporcionados, se espera que realices un análisis exploratorio de los datos para identificar patrones, tendencias y posibles problemas.
- **Preprocesamiento y limpieza de 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 features:** Generamos características basadas en estos datos. Por ejemplo, podríamos resumir la severidad de los síntomas durante los últimos 7 días, el número de veces que el paciente ha mencionado sentirse deprimido, o el número de noches que han reportado tener problemas para dormir.
- **Modelado:** De armar un posible target realiza un modelo de ML de manera libre y explicando las decisiones que tomas para cada paso.
- **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 detendrías 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 hasta el miércoles 12 de julio a las 11:59 pm.

Nota: No esperamos (ni tenemos) una respuesta correcta, preferimos que nos expliques el razonamiento detrás de tus decisiones y si es que hay algo que no se puede realizar que también este explicado sea en texto o a través de código. Diviértete!

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