**Validación manual por muestreo aleatorio**

Consiste en seleccionar una muestra representativa de la población de datos, de forma razonablemente aleatoria y de tamaño suficientemente grande (habitualmente entre el 20% y 40%).

En este caso, la población total consta de 479 preguntas detectadas automáticamente. Se seleccionó una muestra de 160 preguntas, lo que representa el 33,40% del total (160 /479\*100).

Para realizar la selección, se utilizó un intervalo sistemático de 3 para cada conjunto de preguntas obtenidas de del análisis de cada uno de los tres archivos JSON resultados de la exportación de datos de Discord de las conversaciones del año 2024.

La revisión se hizo manualmente, con criterios claros y previamente definidos para evaluar tanto las preguntas como las respuestas asociadas.

**Criterios de evaluación de detección de preguntas**

|  |  |  |  |
| --- | --- | --- | --- |
| Tipo de evaluación | Puntos | Descripción | Ejemplo |
| SI | 1 | El mensaje es efectivamente una pregunta (académica o administrativa). | “Buenas! ¿Qué tal? Quería saber si iban a subir el código de lo que hizo Eze hoy en clase.”  “¿La única forma de hacerlo es creando una lista desde antes?” |
| NO | 0 | El mensaje no es una pregunta, sino un agradecimiento, afirmación o fragmento de conversación. | “Sí, eso también va en el final.” |

**Criterios de evaluación de respuestas asociadas**

|  |  |  |  |
| --- | --- | --- | --- |
| Tipo de evaluación | Puntos | Descripción | Ejemplo |
| SI | 1 | Al menos una de las respuestas detectadas responde correctamente el contenido principal de la pregunta. Se acepta que haya respuestas adicionales irrelevantes. | Una pregunta con tres respuestas, pero una responde adecuadamente. |
| NO | 0 | La pregunta no tiene respuesta, o las respuestas detectadas no contestan la pregunta. | A una consulta técnica sobre Git se responde solo con “Gracias, era eso!”. |

**Métricas de calidad evaluadas**

En esta etapa se calculó la precisión en la detección respuestas:

* + Precisión de respuestas =

Donde "respuestas correctas" hace referencia a aquellas que efectivamente responden a la pregunta detectada. Esta evaluación se realiza por hilo de conversación, no por mensaje individual. Se considera una respuesta correcta si contesta a la pregunta, aunque haya otras respuestas irrelevantes alrededor.

No se calculó recall ya que no se conoce la cantidad total real de preguntas o respuestas existentes en los archivos JSON.

* + - * Recall =

F1-score tampoco puede calcularse porque requiere combinar precisión y recall:

* + - * F1-score =

Resultados

La validación se realizó el 26 de MAYO de 2025, sobre los archivos:

* “ log\_preguntas\_efectivamente\_persistidas\_1”
* “ log\_preguntas\_efectivamente\_persistidas\_2”
* “ log\_preguntas\_efectivamente\_persistidas\_3”
* “ log\_procesamiento\_con\_preguntas\_cerradas”

donde se registraron las preguntas y respuestas detectadas automáticamente.

* + Cantidad de preguntas evaluadas: 160
    - Del primer lote se analizaron 63 preguntas, me dio 90,48% respondidas. Como el margen de error es ±7.25%, el valor real podría estar entre 83,23 % y 97,73%.
    - En el segundo analicé 51 preguntas, me dio 92% respondidas. Como el margen de error es ±7.38%, el valor real podría estar entre 84,62 % y 99,38%.
    - En el tercer lote analicé 46 preguntas, me dio 85% respondidas. Como el margen de error es ±10.38%, el valor real podría estar entre 74,62 % y 95,38%.

Tendencias observadas

* Buena precisión en respuestas: aunque algunas preguntas quedaron sin respuesta o fueron parcialmente respondidas, se justifica por la naturaleza de las conversaciones en Discord: hay idas y vueltas de mensajes de distintos usuarios.