

Implementación de un modelo de traducción de lenguaje natural para la consulta de datos de Google Analytics 4

GitHub
<https://github.com/danielurrutxua/NL-to-GA4-Query>

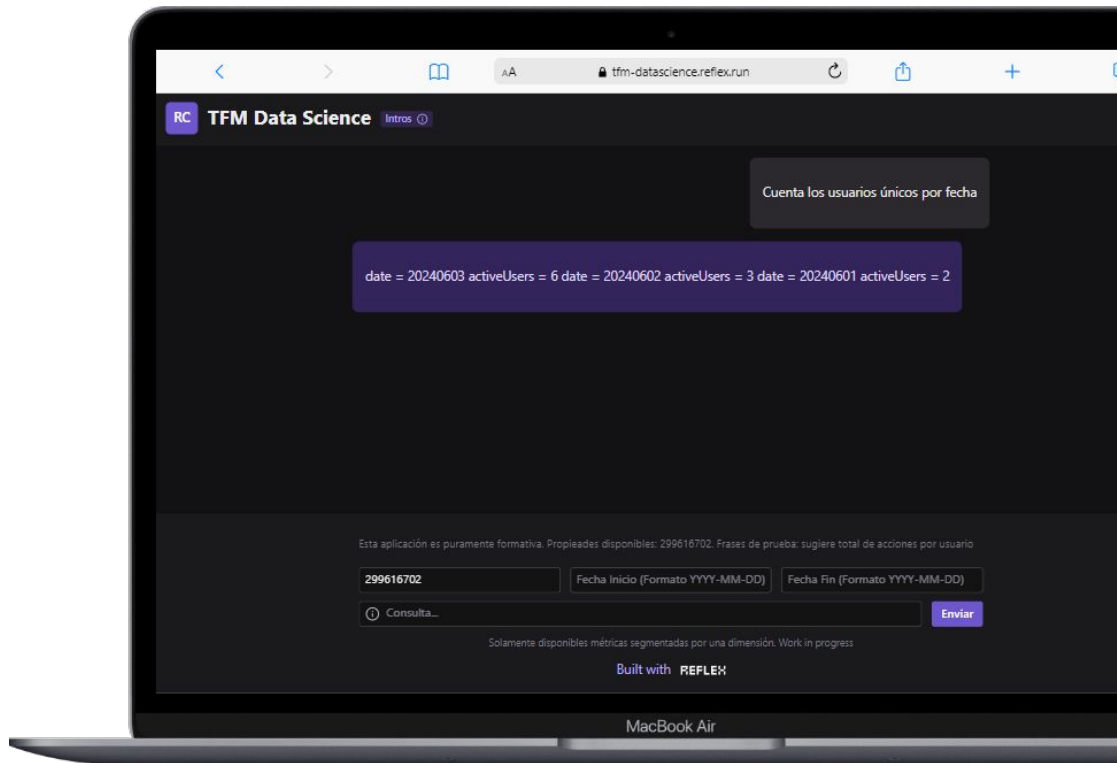
Alberto Sebastián
Álvaro Salmerón
Daniel Urruchua
Juan Ignacio Alberola

Contexto

Este proyecto desarrolla un modelo de **traducción de lenguaje natural** para consultas en Google Analytics 4 (GA4), permitiendo a usuarios sin conocimientos técnicos acceder a datos valiosos.

Utilizando **técnicas avanzadas de procesamiento de lenguaje natural (NLP)** y desplegado en Google Cloud, la herramienta traduce consultas en lenguaje natural a métricas y dimensiones de GA4, optimizando la toma de decisiones y reduciendo la dependencia de analistas de datos.

A futuro, se planea **mejorar** el modelo, **expandir** sus funcionalidades y comercializar la herramienta mediante **suscripciones**.



Las herramientas en la empresa

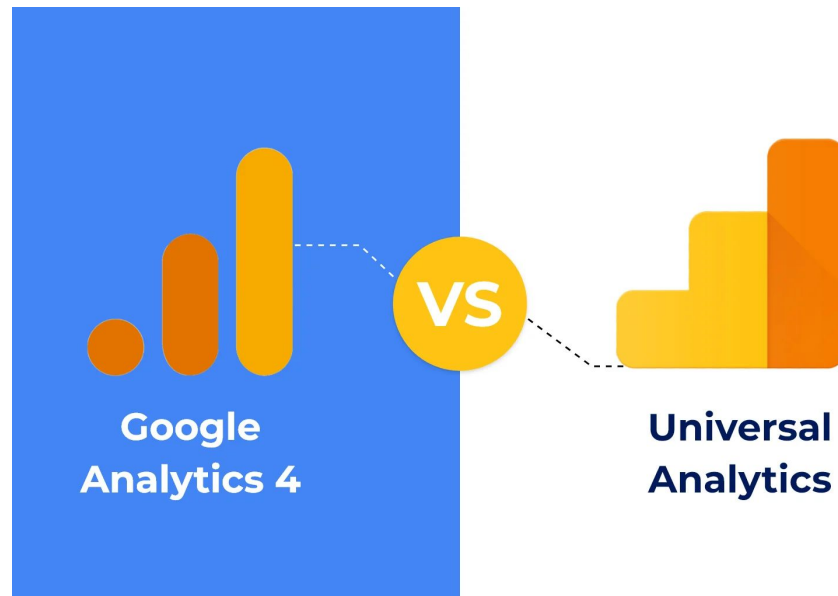
El equipo de marketing de la empresa ha utilizado históricamente Universal Analytics para analizar y comprender el comportamiento de los usuarios en sus sitios web y aplicaciones móviles. Debido a la actualización a Google Analytics 4 (GA4), han adoptado esta nueva versión para aprovechar sus capacidades avanzadas y mejorar el análisis de datos.

Ventajas

- **Análisis Completo:** Visión amplia y detallada de la interacción de los usuarios.
- **Modelo Basado en Eventos:** Mayor flexibilidad y profundidad en el análisis de datos.

Desventajas

- **Curva de Aprendizaje:** Requiere mayor conocimiento y comprensión de la herramienta.



La problemática en el negocio

Nos encontramos con un entorno en el que los comerciales dependen del equipo de analistas, restando así su capacidad para atender más leads mensualmente. Buscamos:

Accesibilidad a los datos

Permitir a usuarios no técnicos acceder a datos de Google Analytics 4 (GA4) mediante consultas en lenguaje natural

Reducir la dependencia de equipos técnicos especializados

Toma de decisiones

Facilitar la obtención de insights para la toma de decisiones estratégicas y operativas

Optimización de recursos

Liberar a los analistas de datos de consultas rutinarias para que se centren en tareas más complejas

Un ejemplo económico

En una empresa con **5 comerciales** en la situación actual donde deben solicitar los datos al equipo de analistas **cada uno de ellos puede atender 8 leads al mes.**

Cada contrato cerrado tiene un **ticket medio de 10.000€** y la tasa de conversión, tras revisar el histórico de la empresa es del 15%.

Mediante entrevistas con el departamento comercial se ha llegado a la conclusión de que si tienen la capacidad de consultar datos básicos sobre los clientes sin esperar a las respuestas de los analistas, **podrían atender a 10 leads mensuales.**

El incremento de facturación anual se ha determinado importante para afrontar la escalabilidad de la empresa para contratar a más equipo.

Sin la herramienta

Ventas mensuales

$$\begin{aligned} &5 \text{ comerciales} * 8 \text{ leads} * 15\% \text{ CR} = 6 \text{ ventas} \\ &\text{Ingresos} = 6 \text{ ventas} * 10\text{k} = 60.000\text{€ mensuales} \\ &\text{Anual} = 60.000 * 12 = \mathbf{720.000\text{€}} \end{aligned}$$

+180k€
Anuales

Con la herramienta

Ventas mensuales

$$\begin{aligned} &5 \text{ comerciales} * 10 \text{ leads} * 15\% \text{ CR} = 7,5 \text{ ventas} \\ &\text{Ingresos} = 7,5 \text{ ventas} * 10\text{k} = 75.000\text{€ mensuales} \\ &\text{Anual} = 75.000 * 12 = \mathbf{900.000\text{€}} \end{aligned}$$

¿Cómo lo queremos conseguir?

Ofrecer una solución interna que permita generar más ingresos para reinvertir en otros departamentos de la empresa para poder escalarla. Posteriormente disponer de un producto que comercializar de forma externa.

Desarrollo modelo NLP

Crear un modelo que traduce lenguaje natural en métricas y dimensiones de GA4

Despliegue en producción

Implementar la solución en Google Cloud para asegurar su escalabilidad y disponibilidad

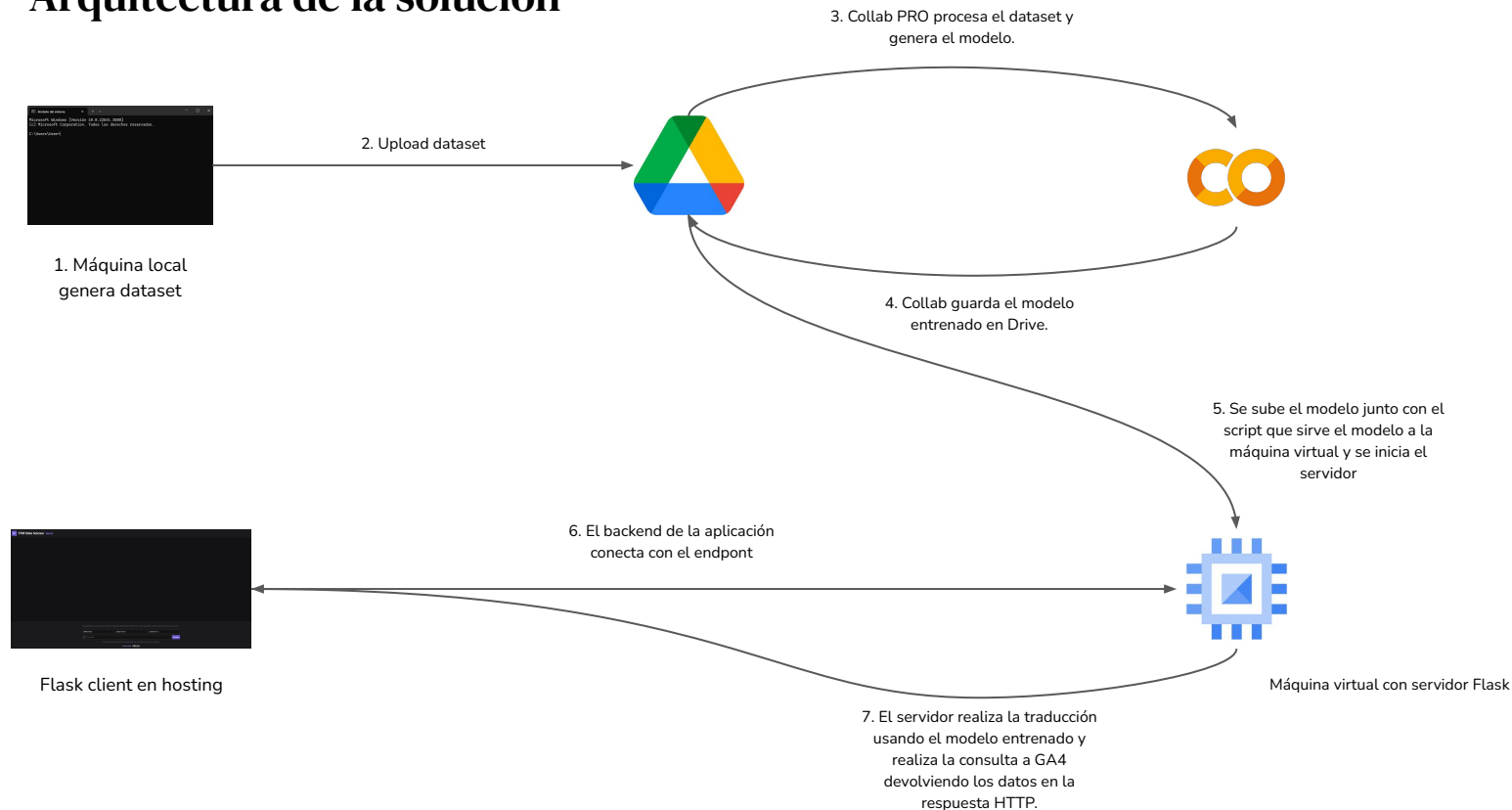
Interfaz de usuario

Diseñar una interfaz que permita a los usuarios formular preguntas fácilmente

¿Qué proceso se ha seguido?



Arquitectura de la solución



Puesta en valor

RC

TFM Data Science

Intros ⓘ

Esta aplicación es puramente formativa. Propiedades disponibles: 299616702. Frases de prueba: sugiere total de acciones por usuario

299616702

2024-05-01

2024-05-12

ⓘ Consulta...

Enviar

Solamente disponibles métricas segmentadas por una dimensión. Work in progress

Built with REFLEX

La puesta en valor inicial se ha realizado a través de una aplicación web donde los comerciales pueden realizar las consultas.

No obstante, con el futuro desarrollo del sistema se puede llegar a integrar con equipos de marketing y gerentes.

Acciones futuras

Mejora continua del modelo

Incorporar consultas erróneas para entrenar el modelo y aumentar su **precisión**.

Expansión de funcionalidades

Desarrollar el modelo para comprender una mayor **variedad de consultas temporales, métricas y dimensiones**

Monetización

Implementar un modelo de **suscripción y explorar alianzas estratégicas**

LLMs

Hacer uso de las múltiples APIs de las LLMs para **plantear una alternativa tecnológica** a la solución propuesta.

Problemática encontrada

En el transcurso del desarrollo de este proyecto nos hemos encontrado con tres dificultades principales.

Variedad del lenguaje

Ha resultado complejo el poder capturar una mínima parte de las maneras de realizar solicitudes al modelo para la variedad de métricas y dimensiones existentes en GA4.

Capacidad computacional

Al necesitar datasets ciertamente grandes para poder capturar la mayor variedad del lenguaje en nuestras máquinas locales ha sido complicado el poder entrenar el modelo por lo que se ha usado Google Collab PRO.

Evaluación del modelo

Para ilustrar la evaluación del modelo se ha tenido que usar un dataset más pequeño debido al tiempo necesario para cada prueba de entrenamiento en el ajuste de hiper parámetros.



GitHub

<https://github.com/danielurrutxua/NL-to-GA4-Query>

Alberto Sebastián

[\(albertosebastianpascual@gmail.com\)](mailto:albertosebastianpascual@gmail.com)

Álvaro Salmerón

[\(alvarosh31@gmail.com\)](mailto:alvarosh31@gmail.com)

Daniel Urruchua

[\(durrutxua@gmail.com\)](mailto:durrutxua@gmail.com)

Juan Ignacio Alberola

[\(juanignacioalberola@gmail.com\)](mailto:juanignacioalberola@gmail.com)