

Proyecto 1 – Etapa 2: Analítica de Textos

Universidad de los Andes

Departamento de Ingeniería de Sistemas y Computación

Inteligencia de Negocios

Santiago Páez Márquez - 2020214644

Luis Ángel Angarita Coba – 201910393

David Mateo Barbosa Monsalve - 202110756

Proceso de automatización del proceso de preparación de datos, construcción del modelo y acceso a la aplicación	2
Desarrollo de la aplicación y justificación	3
Resultados	5

Proceso de automatización del proceso de preparación de datos, construcción del modelo y acceso a la aplicación

Para automatizar el proceso de preparación de datos y la construcción del modelo de análisis de texto, se desarrolló un pipeline que integró la metodología de preprocesamiento, como tokenización, eliminación de stopwords, lematización y vectorización TF-IDF. Este pipeline, basado en las funciones implementadas en la primera fase del proyecto, garantiza que los datos pasen por los mismos pasos antes de preparar el modelo. Una vez completado, el pipeline fue exportado como un archivo .joblib, lo que permite a la organización tener una versión persistente del modelo sin necesidad de repetir el proceso desde cero.

Para hacer accesible este modelo y su pipeline a través de una aplicación web, se utilizó Flask junto con la biblioteca request, proporcionando una interfaz de programación de aplicaciones API. Flask se seleccionó por su simplicidad y flexibilidad, permitiendo el desarrollo ágil de la aplicación web. A través de esta API, los usuarios pueden enviar solicitudes HTTP que desencadenan la carga del archivo .joblib y la ejecución del pipeline en tiempo real.

El uso de Flask y request facilita la conexión entre el frontend de la aplicación web y el modelo analítico. La API recibe los datos de texto, los procesa utilizando el pipeline preentrenado y devuelve resultados de manera eficiente. Esta arquitectura, al ser ligera y de fácil mantenimiento, permite una integración rápida con cualquier frontend que necesite interactuar con el modelo.

Desarrollo de la aplicación y justificación

Para la consolidación del desarrollo de la aplicación, se tuvieron en cuenta diferentes tipos de usuario que podrían hacer uso de la aplicación:

Rol dentro de la organización	Tipo de actor	Beneficio	Riesgo
Dirección de Políticas Públicas	Usuario-cliente	Facilita la identificación de problemas clave en salud, educación y género, permitiendo crear políticas más eficientes y enfocadas en áreas prioritarias.	Si el modelo tiene un mal desempeño, podría priorizar problemas irrelevantes y descuidar cuestiones urgentes.
Fondo de Poblaciones de las	Financiador	Reduce costos operativos	Si el modelo no alcanza la

Naciones Unidas (UNFPA)		mediante la automatización del análisis, mejorando la eficiencia en la toma de decisiones basadas en datos.	precisión esperada, los recursos invertidos podrían no generar el impacto esperado.
Ministerio de Tecnologías de la Información y las comunicaciones	Proveedor	Asegura el cumplimiento de estándares de calidad, seguridad y privacidad en el manejo de datos ciudadanos durante el desarrollo del modelo.	El manejo incorrecto o la filtración de datos podría generar problemas legales y pérdida de confianza pública.
Ciudadanos	Beneficiado	Sus problemas relacionados con salud, educación y género pueden ser abordados más rápidamente y con mayor precisión gracias a las políticas públicas generadas a partir del modelo.	Si el modelo no funciona adecuadamente, podría ignorar problemas críticos en las comunidades más vulnerables

Con el propósito de cumplir los objetivos planteados, se decidió que la aplicación debía estar dirigida principalmente a usuarios del Fondo de Poblaciones de las Naciones Unidas (UNFPA). Esta herramienta permite a sus usuarios ingresar nuevos testimonios de ciudadanos y obtener predicciones sobre a cuál Objetivo de Desarrollo Sostenible (ODS) están relacionados, basándose en un modelo analítico desarrollado.

Este enfoque facilita el análisis automatizado de las opiniones y experiencias de los ciudadanos, que reflejan problemáticas locales en torno a la salud, la educación y el género. Al utilizar la aplicación para procesar estos testimonios, se optimiza el uso de los recursos, ya que el tiempo y el esfuerzo previamente dedicados a analizar manualmente los datos pueden ahora enfocarse directamente en implementar acciones dentro de la agenda de los ODS. El resultado es un enfoque más eficiente en la asignación de recursos, logrando que las voces de las personas sean rápidamente conectadas con los objetivos de desarrollo.

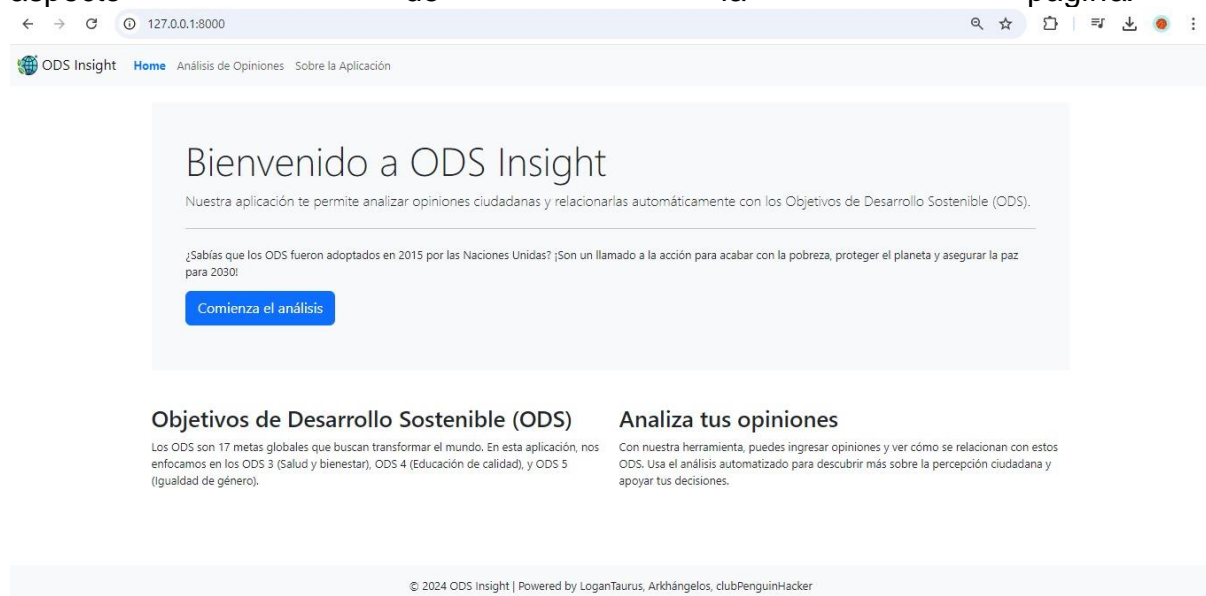
El proyecto no hubiera sido posible sin la colaboración clave de actores como el Ministerio de Tecnologías de la Información y las Comunicaciones, que garantizó que el manejo de los datos ciudadanos cumpliera con los más altos estándares de seguridad y privacidad. Asimismo, la Dirección de Políticas Públicas participó activamente en el desarrollo, asegurando que el modelo fuera una herramienta útil para la creación de políticas más enfocadas, que respondieran a las necesidades reales de las comunidades.

El trabajo entre las diferentes áreas con el grupo de estadística fue crucial. Se concluyó que la mejor manera de lograr los resultados esperados era automatizar por completo el análisis de texto y ajustar el modelo para que fuera fácil de usar por los diferentes actores involucrados, incluidos los usuarios de la UNFPA. Esta simplicidad en el uso no compromete la efectividad del modelo, permitiendo a cualquier usuario interactuar de manera intuitiva con la plataforma y obtener resultados relevantes de manera rápida.

Finalmente, el equipo de estadística aportó sus conocimientos para definir cómo se debía construir el enfoque analítico, evitando la sobrecarga de información técnica y facilitando el acceso a los resultados sin necesidad de presentar métricas matemáticas complejas. Esto aseguró que la aplicación fuera accesible no solo para expertos, sino para cualquier persona involucrada en la toma de decisiones, desde el financista hasta los ciudadanos beneficiados. Como resultado, el desarrollo no solo ofrece precisión analítica, sino también una experiencia de usuario sencilla y directa, clave para el éxito del proyecto.

Resultados


Toda la descripción y el análisis de la aplicación se encuentra en el video, sin embargo, se va a mostrar una pequeña muestra de las funcionalidades y aspecto de la página.



La imagen anterior muestra la pantalla de inicio donde habla de los objetivos de

desarrollo sostenible y cuales nos enfocaremos, la funcionalidad de la aplicación y un botón para iniciar el análisis.

Esta aplicación tiene dos funcionalidades, en primer lugar, tenemos el entrenamiento de los datos donde únicamente se debe subir el archivo con los datos a entrenar.

 ODS Insight

[Home](#) [Análisis de Opiniones](#) [Sobre la Aplicación](#)

Cargar Datos de Entrenamiento

Sube un archivo con datos de entrenamiento en formato CSV, Excel o JSON.

Entrenamiento del modelo completado con éxito

Modo avanzado

Archivo de entrenamiento (CSV, Excel o JSON):

Choose File

No file chosen

Subir Archivo

Resultados del Entrenamiento

Métrica	Valor
Precisión	98.23%
Recall	98.22%
F1-Score	98.22%

Regresar al Analizador

© 2024 ODS Insight | Powered by LoganTaurus, Arkhángelos, clubPenguinHacker

En

segundo lugar, tenemos el analizador de textos los cuales, con un modelo entrenado, realiza la clasificación de textos a la ODS correspondiente con su probabilidad. Esta opción permite subir un archivo de todas las opiniones para un análisis conjunto.

Análisis de Opiniones

Analiza opiniones ingresadas o sube un archivo con datos en formato CSV, Excel o JSON para analizarlas automáticamente.

Predicción exitosa

Ingresar Opinión

Subir Archivo

☒ Modo avanzado

Carga un archivo (CSV, Excel o JSON):

Choose File

No file chosen

Analizar Archivo

Resultados del Análisis

Texto	SDG	Probabilidad
Han examinado la contribución de las universidades y las instituciones de educación terciaria al desarrollo del capital humano y las competencias; la transferencia de tecnología y la innovación empresarial, el desarrollo social, cultural y medioambiental, y la creación de capacidad regional. El proceso de revisión ha facilitado la creación de asociaciones en ciudades y regiones al reunir a instituciones de educación terciaria y organismos públicos y privados para identificar objetivos estratégicos y trabajar juntos para alcanzarlos. Complementa las revisiones que se han llevado a cabo en la región transfronteriza México-Estados Unidos, de gran importancia estratégica y económica, como las de Nuevo León, la región Paso del Norte y, más recientemente, la del sur de Arizona (2011), junto al estado de Sonora.	4	76.0%
En la última década, y en particular desde 2010, el número de altas per cápita ha tendido a disminuir, y ahora está en línea con la media de la OCDE mencionada anteriormente. Por ejemplo, los habitantes de zonas rurales representan el 43% de la población, pero sólo el 32% de las altas hospitalarias. A falta de información complementaria, no es posible determinar si esta diferencia	3	92.0%

También se puede realizar el análisis de cada una de las opiniones por separado, solo se debe colocar en el cuadro de texto la opinión y esta devolverá el ODS correspondiente y su probabilidad.

Análisis de Opiniones

Analiza opiniones ingresadas o sube un archivo con datos en formato CSV, Excel o JSON para analizarlas automáticamente.

Ingresar Opinión

Subir Archivo

☐ Modo avanzado

Escribe tu opinión:

Escribe aquí tu opinión...

Analizar Opinión

Resultados del Análisis

Texto	SDG
Por ejemplo, el número de consultas externas de especialistas es de 319 por cada mil derechohabientes en el SP, en comparación con 338 y 620 por cada mil derechohabientes en el IMSS y el ISSSTE, respectivamente. Si bien algunas de estas diferencias pueden reflejar una necesidad desigual (como la población ligeramente mayor del ISSSTE), otras no pueden justificarse de esta manera. El número de recetas que no pudieron ser surtidas en su totalidad por un farmacéutico debido a la falta de existencias es de 33% dentro del SP en comparación con 14% dentro del IMSS según los datos de la encuesta (aunque las propias cifras de los institutos de la SS sugieren tasas más altas de recetas surtidas). Ambas cifras se encuentran entre las más altas de la OCDE. El gasto de bolsillo no se ha reducido significativamente en la última década, a pesar de los esfuerzos para lograr la cobertura sanitaria universal a través de la reforma del SP.	5

Finalmente, después de obtener los resultados del análisis, hay una opción de descargar los resultados para posibles usos posteriores.

Descargar Resultados

- ☐ CSV
- ☒ Excel
- ☐ JSON

Descargar

A grandes rasgos, se evidencia que la aplicación tiene una interfaz amigable con el usuario y unos botones fáciles de interpretar para que cualquier persona pueda interactuar con la página.

Estudiantes	Rol	Horas trabajadas	Tareas hechas	Puntaje
Luis Ángel	Líder de proyecto – Ingeniero de Datos	15	Automatización para el Pipeline y diseño del Back	32
Santiago Páez	Ingeniero de software responsable de diseño y resultados	12	Documento – Vídeo Resultados	32
David Mateo	Ingeniero de Software responsable de la aplicación final	16	Diseño de la aplicación web	36