

10 de abril de 2024

Actividad: Creación del grupo en Discord y organización inicial

- Se creó un grupo en Discord para facilitar la comunicación entre los miembros del proyecto.
- Se mencionó el uso de herramientas como Tableau para el análisis de datos y se discutió sobre la configuración de cuentas y permisos para acceder a los recursos compartidos.

Actividad: Enlace compartido y configuración inicial

- Se compartieron enlaces relevantes, incluyendo el dataset de votaciones de la ONU y algunos artículos relacionados sobre análisis de bloques en las votaciones.
- Se decidió iniciar el trabajo con una reunión al día siguiente para definir los primeros pasos y comenzar a trabajar en los datos.

11 de abril de 2024

Actividad: Subida de Jupyter Notebook y discusión inicial

- Se subió un Jupyter Notebook con un modelo preliminar script de web scraping que descargaba datos de votaciones y los procesaba para una visualización inicial.
- Se discutieron problemas potenciales con la cantidad de datos disponibles y se propuso ampliar el alcance del análisis, revisando y ajustando el script de filtrado para asegurar que todos los datos relevantes fueran incluidos.
- Se estableció un repositorio en GitHub y se agregaron los miembros del equipo al mismo.

12 de abril de 2024

Actividad: Coordinación de tareas

- Se realizó una reunión en Google Meet para discutir el progreso y asignar nuevas tareas.

13 de abril de 2024

Actividad: Revisión de scripts y tareas asignadas

- Se revisó el script de filtrado y el conjunto de datos disponibles, identificando la necesidad de categorizar temas más grandes como cambio climático y derechos humanos para aumentar la variabilidad del análisis. Luego se determinó que no era necesario.
- Se discutió la alineación de objetivos y posibles ajustes a los datos del proyecto, decidiendo ajustar las URLs del código de web scraping para incluir temas más amplios y mejorar los gráficos generados.

14 de abril de 2024

Actividad: Descubrimiento de nuevos datasets

- Encontramos un dataset en Hugging Face que contenía datos de votaciones de la ONU desde 2000 hasta 2023.
- Se compartió este recurso con el equipo y comenzó a trabajar en la integración de estos datos con la base existente para un análisis más completo .

15 de abril de 2024

Actividad: Preparación de datos y análisis exploratorio

- Se propuso preparar el DataFrame (df) con los datos actualizados y verificar la agrupación de países basada en el Consejo de Seguridad.
- También comenzamos con el análisis exploratorio en Tableau y con el diagrama de la arquitectura.

16 de abril de 2024

Actividad: Diagrama de arquitectura y preparación de datos.

- Se subió el diagrama de arquitectura al repositorio.

17 de abril de 2024

Actividad: Coordinación y reparto de tareas

- Se revisaron los puntos de la práctica juntos y se discutió el valor de los datos.

18 de abril de 2024

Actividad: Reunión de planificación

- Se discutió el progreso de la preparación de datos y se planificaron las siguientes tareas a realizar, asignando responsabilidades a cada miembro del equipo .

19 de abril de 2024

Actividad: Análisis de datos y preparación para el modelo

- Se continuó trabajando en la preparación de los datos y en la configuración del entorno para el entrenamiento del modelo.

20 de abril de 2024

Actividad: Finalización de la preparación de datos

- Se completó la preparación del DataFrame y se realizaron pruebas iniciales con los datos para asegurar que estaban listos para el análisis.

26 de abril de 2024

Actividad: Actualización y estructuración de datos

- Se comentó sobre la estructuración de datos para el procesamiento de lenguaje natural (NLP) y la digitalización de resoluciones de la ONU.
- Se mencionó la disponibilidad de la base de datos de Harvard con información sobre votaciones de la Asamblea General de la ONU, enfocándose en el conflicto de Palestina.

29 de abril de 2024

Actividad: Reunión y carga de archivos

- Compartimos los correos electrónicos de los miembros para agendar una reunión con Cristina. Confirmaron la reunión para el 1 de mayo a las 15 horas (hora de España).
- Se cargó un archivo de Tableau actualizado con la vista consolidada y se pidieron comentarios para realizar ajustes .

30 de abril de 2024

Actividad: Coordinación de reunión

- Se discutieron las alternativas de horarios para la reunión con Cristina y se decidió que el viernes 2 de mayo era el mejor día para la mayoría .

3 de mayo de 2024

Actividad: Preparación y reunión

- Se coordinó una reunión para revisar juntos lo que se iba a discutir con Cristina.
- Después hablamos de cómo organizarnos el trabajo de cara al fin de semana.
- Se subió el notebook realizado de conversión de archivos.

4 de mayo de 2024

Actividad: Procesamiento de documentos y preparación de modelos

- Se continuó con el procesamiento de documentos y la preparación de modelos para análisis de datos, utilizando las herramientas y scripts desarrollados previamente .

5 de mayo de 2024

Actividad: Coordinación y asignación de tareas

- Se coordinó una reunión para asignar tareas específicas a cada miembro del equipo y asegurar que todos estuvieran alineados con los objetivos del proyecto .

6 de mayo de 2024

Actividad: Coordinación de reuniones

- Se propuso coordinar una reunión para discutir la disponibilidad y planificar las actividades de los próximos días. Se acordó una llamada en Google Meet para asegurar que todos los miembros estuvieran alineados .
- Se cargó el flujo de datos recomendado por Cristina en la reunión.
- Se realizaron ajustes al documento "Análisis_Exploratorio_UNVote" en Tableau, específicamente en la gráfica de distribución de votaciones de los países miembros del Consejo de Seguridad con respecto a resoluciones relacionadas a opiniones consultivas enviadas a la CIJ.
- También se trabajó en el diagrama de flujo del proceso que abarca desde la votación en la Asamblea General hasta el envío de la opinión consultiva a la CIJ, pasando por la generación de insumos y el retorno a la Asamblea General para la votación sobre la adopción de medidas sugeridas por la Corte Internacional de Justicia.
- Además, se llevó a cabo la reunión de coordinación donde se presentaron los avances hasta la fecha en temas de visualización de datos, procesamiento, y la implementación de webscraping y crawling, aspectos fundamentales para el análisis de sentimientos.

8 de mayo de 2024

Actividad: Reunión

- Nos conectamos para discutir sobre la conversión de archivos a un formato JSON, buscando diferentes herramientas como PyPDF2

9 de mayo de 2024

Actividad: Limpieza de archivos

- Se informó sobre la creación de una nueva carpeta llamada "Archivos Segmentados".
- También se mencionó la limpieza de estos archivos para prepararlos para el procesamiento.
- Se desarrolló un script empleando PyPDF2 y Pandas con el propósito de segmentar las respuestas proporcionadas por la corte en las opiniones consultivas. Estas segmentaciones se almacenan en archivos CSV individuales, los cuales se guardan en la carpeta "ArchivosSegmentados". Posteriormente, estos archivos serán sometidos a

un proceso de eliminación de párrafos superfluos, así como la inclusión de la conclusión de la corte en relación al tema tratado y la pregunta planteada.

10 de mayo de 2024

Actividad: Limpieza de datos y preparación de archivos

- Se trabajó en la limpieza de archivos CSV y en la preparación de datos segmentados para su procesamiento, asegurando que estuvieran listos para el análisis .

13 de mayo de 2024

Actividad: Entrenamiento de modelos y creación de JSON

- Se comenzó con la preparación del entorno para el entrenamiento del modelo utilizando una GPU local.
- También se trabajó en la creación de archivos JSON de opiniones consultivas y otros documentos relevantes.
- Logramos montar el JSON de la primera pregunta de Advisor Opinions. Ahora necesitamos hacer lo mismo para el resto de las preguntas.
- Queremos dividir el texto de los Oral Proceedings en JSON para clasificarlo según su contenido.
- El archivo de conversión se ha actualizado para validar el contenido del dataset.

15 de mayo de 2024

Actividad: Avances en procesamiento de documentos

- Se consolidaron los datos procesados y se actualizaron los archivos en el repositorio de GitHub, asegurando que todos los miembros del equipo tuvieran acceso a la versión más reciente de los datos y scripts.

16 de mayo de 2024

Actividad: Finalización de documentos y organización de tareas

- Se finalizó el procesamiento de archivos orales y se subieron los JSON al repositorio.
- Se destacó la importancia de adelantar el trabajo con los Oral 2004 para completar el análisis .

17 de mayo de 2024

Actividad: Avances en procesamiento de archivos

- Se continuó con el procesado de los archivos Verbatim record 2024 y se mencionaron avances en la validación del modelo.
- Se identificaron problemas con el sobreajuste del modelo y se discutió la posible necesidad de ajustar el enfoque.

18 de mayo de 2024

Actividad: Validación del modelo y ajustes adicionales

- Se realizaron pruebas para validar el modelo NLP y se discutieron posibles ajustes para mejorar su desempeño y evitar el sobreajuste.
- Hubo dificultades a la hora de configurar correctamente la GPU en local y se trabajó para resolverlo.

19 de mayo de 2024

Actividad: Avances en procesamiento de archivos

- Se continuó con el procesado de los archivos Verbatim record 2024 y se mencionaron avances en la validación del modelo.
- Se identificaron problemas con el sobreajuste del modelo y se discutió la posible necesidad de ajustar el enfoque.

20 de mayo de 2024

Actividad: Consolidación de datos procesados.

- Se informó sobre la finalización del procesado de todos los Orals de 2024 y se subió el consolidado y los archivos individuales restantes.

- Se realizaron pruebas de entrenamiento del modelo usando la GPU con distintas configuraciones de parámetros. Se organizó una llamada para probar y validar los resultados del modelo entrenado.

21 de mayo de 2024

Actividad: Actualización de bitácora y repaso general

- Se trabajó sobre la bitácora para ponerla al día.
- Se actualizó el diagrama funcional de la estructura.
- Se continuó con alguna prueba más del modelo para tratar de afinarlo lo máximo posible antes de darlo por terminado.
- Se creó un nuevo repositorio en GitHub para ir añadiendo cada parte del proyecto de una forma organizada y limpia.

22 de mayo de 2024

Actividad: Toques finales

- Se actualizaron los readme correspondientes a cada módulo.
- Terminado el notebook del modelo final junto con Q&A correspondiente

23 de mayo de 2024

Actividad: Entrega

- Finalizados los readme de cada módulo
- Hechas correcciones y explicaciones finales a los notebook
- Realizadas diapositivas para presentación final
- Entrega del repositorio en el formulario