

Guía del Proyecto Análisis Exploratorio

INTRODUCCIÓN

Existen muchos problemas complejos que se pueden abordar con la Ciencia de Datos. Para este proyecto se plantean varias alternativas. La actividad se realizará en grupos de 3 a 5 personas que deberán seleccionar uno de los siguientes retos de los que aquí se plantean:

#	Reto	Tema
	CommonLit - Evaluar resúmenes de estudiantes	Procesamiento del Lenguaje Natural
1	Detección de estructuras microvasculares en tejidos de riñón humano sanos.	Visión Artificial
2	Detectar y clasificar lesiones abdominales traumáticas	Visión Artificial
3	Google: reconocimiento de deletreo manual del lenguaje de señas estadounidense	Visión Artificial
4	Desafío Ojos en el Terreno del CGIAR. Detección de enfermedades en plantas	Visión Artificial
5	Identificación de Especies de Mosquitos	Visión Artificial
6	Sistema de recomendación de Spotify	Sistema de recomendación
7	Hackeando el cuerpo humano	Visión Artificial
8	Predicción de argumentos efectivos	Procesamiento del Lenguaje Natural
9	Detección de fracturas de las vértebras cervicales en radiografías	Visión Artificial
10	Detección de pases en videos de jugadas de football de la liga alemana	Visión Artificial
11	Clasificación de los orígenes de un coágulo de sangre en un accidente cerebrovascular	Visión Artificial
12	Identificación de menciones de entidades biomédicas en resúmenes de artículos de investigación	Procesamiento del Lenguaje Natural
13	Identificación de relaciones de entidades biomédicas en resúmenes de artículos de investigación	Procesamiento del Lenguaje Natural

INSTRUCCIONES

Sobre los Grupos:

- Deben inscribirse a alguno de los grupos de Canvas para el proyecto 2. Si no se inscribe en algún grupo no será calificado. **Aunque el Canvas lo permite, no se aceptarán grupos con menos de 3 estudiantes.**

Sobre los Retos:

- Cada grupo debe seleccionar un reto diferente, no pueden repetirse. Si esto sucediera se revisará la entrega del primer grupo que la realice.

Sobre el proyecto:

- Puede usar los recursos que Kaggle o Google Colab le proporciona, pero debe versionarlo en Github puesto que sus contribuciones se usarán para la evaluación individual de cada miembro del grupo.

ACTIVIDADES

1. Haga una pequeña investigación del tema para que tenga idea de qué buscar en un análisis exploratorio. En el caso de los problemas médicos, describa la enfermedad a detectar, los síntomas y como se diagnostica (especialmente diagnóstico basado en imágenes). Esto le va a servir para entender cual es el patrón que deben reconocer los algoritmos. En el caso de problemas de Procesamiento del Lenguaje Natural, investigue las técnicas que se usan para detectar patrones en lenguaje escrito. En cuanto al tema de las recomendaciones investigue que se necesita para hacer un sistema de recomendaciones.
2. Analice el problema planteado y los datos.
3. Describa las tareas de limpieza y preprocesamiento que llevó a cabo.
4. Haga un análisis exploratorio de los datos:
 - a. Comience describiendo cuantas variables y observaciones tiene disponible, el tipo de cada una de las variables.
 - b. Haga un resumen de las variables numéricas y tablas de frecuencia para las variables categóricas, escriba lo que vaya encontrando, si aplica.
 - c. Cruce las variables que considere que son las más importantes para hallar los elementos clave que lo pueden llevar a comprender lo que está causando el problema encontrado.
 - d. Haga gráficos exploratorios que le de ideas del estado de los datos.
5. Escriba unas conclusiones con los hallazgos encontrados durante el análisis exploratorio

EVALUACIÓN

NOTA: La evaluación de cada integrante del grupo será de acuerdo con sus contribuciones al trabajo grupal

- **(10 puntos) Situación Problemática:** Describe la situación problemática que da lugar al problema.
- **(10 puntos). Problema científico:** Se enuncia el problema científico que se desprende de la situación planteada. Se comprende bien cuál es el problema.
- **(10 puntos). Objetivos:** Se plantean los objetivos a cumplir para darle solución al problema planteado. Se enuncia al menos un objetivo general y 2 específicos. Los objetivos deben ser medibles y alcanzables durante la investigación.
- **(20 puntos). Descripción de los datos:** Se describen los datos, tanto las variables y observaciones como las operaciones de limpieza que se le hicieron si fueron necesarias.
- **(30 puntos). Análisis Exploratorio:**
 - o Estudia las variables cuantitativas mediante técnicas de estadística descriptiva

- Hace gráficos exploratorios como histogramas, diagramas de cajas y bigotes, gráficos de dispersión que ayudan a explicar los datos
- Analiza las correlaciones entre las variables, trata de explicar los outliers (puntos atípicos) y toma decisiones acertadas ante la presencia de valores faltantes.
- Estudia las variables categóricas
- Elabora gráficos de barra, tablas de frecuencia y de proporciones
- Explica muy bien todos los procedimientos y los hallazgos que va haciendo.
- **(20 puntos). Hallazgos y conclusiones:**
 - Hace un resumen de los hallazgos en el análisis exploratorio
 - Llega a conclusiones sobre los siguientes pasos a seguir.

MATERIAL A ENTREGAR

- Archivo .pdf con el informe de análisis exploratorio.
- Link del repositorio usado para versionar el código y/o link de Kaggle o Google Colab en caso de que los utilice.
- Presentación de Power Point a usar para presentar resultados.

FECHAS DE ENTREGA

- **PRESENTACIÓN Y DOCUMENTO FINAL COMPLETO: 19 de septiembre de 2021**

REFERENCIAS

- <https://www.kaggle.com/general/33266>
- <https://medium.com/analytics-vidhya/how-to-use-google-colab-with-github-via-google-drive-68efb23a42d>
-