

Guía para de Proyecto 2. Resultados Finales

INTRODUCCIÓN

La Ciencia de Datos permite abordar una gran variedad de problemas complejos. Con el fin de que los estudiantes desarrollen habilidades prácticas, este año se proponen diferentes retos de proyecto. La actividad se realizará en grupos de 4 integrantes, quienes deberán seleccionar uno de los siguientes desafíos:

#	Reto	Tema
1	Jigsaw - Agile Community Rules Classification	Procesamiento del Lenguaje Natural
2	MITSUI&CO. Commodity Prediction Challenge	Series de tiempo
3	MAP - Charting Student Math Misunderstandings	Procesamiento del Lenguaje Natural
4	RSNA Intracranial Aneurysm Detection	Visión Artificial
5	NeurIPS - Desafío de datos Ariel 2024	Visión Artificial
6	Clasificación degenerativa de la columna lumbar RSNA 2024	Visión Artificial
7	Predecir qué respuesta preferirá un usuario en esta batalla cara a cara con datos del Chatbot Arena	Procesamiento del Lenguaje Natural
8	Predicción de compradores recurrentes: cuestionar la línea base	Negocios
9	CommonLit - Evaluar resúmenes de estudiantes	Procesamiento del Lenguaje Natural
10	Detección de estructuras microvasculares en tejidos de riñón humano sanos.	Visión Artificial
11	Detectar y clasificar lesiones abdominales traumáticas	Visión Artificial
12	Google: reconocimiento de deletreo manual del lenguaje de señas estadounidense	Visión Artificial
13	Desafío Ojos en el Terreno del CGIAR. Detección de enfermedades en plantas	Visión Artificial
14	Identificación de Especies de Mosquitos	Visión Artificial
15	Hackeando el cuerpo humano	Visión Artificial
16	Predicción de argumentos efectivos	Procesamiento del Lenguaje Natural
17	Detección de fracturas de las vértebras cervicales en radiografías	Visión Artificial
18	Detección de pases en videos de jugadas de football de la liga alemana	Visión Artificial
19	Clasificación de los orígenes de un coágulo de sangre en un accidente cerebrovascular	Visión Artificial
20	Identificación de menciones de entidades biomédicas en resúmenes de artículos de investigación	Procesamiento del Lenguaje Natural
21	Identificación de relaciones de entidades biomédicas en resúmenes de artículos de investigación	Procesamiento del Lenguaje Natural

INSTRUCCIONES

Sobre los Grupos:

- Los estudiantes deben inscribirse en los grupos de Canvas para el Proyecto 2. **Si no se inscriben, no recibirán calificación.**

Sobre los Retos:

- Cada grupo debe seleccionar un reto **diferente**. En caso de repetición, se evaluará únicamente la entrega del primer grupo inscrito.

Sobre el proyecto:

- Pueden trabajar en **Google Colab** o **Kaggle**, pero es obligatorio versionar el código en **GitHub**, ya que se evaluarán las contribuciones individuales.

ACTIVIDADES

Aplicación:

1. Haga una aplicación en la que se puedan ingresar nuevos datos y el sistema clasifique usando los modelos entrenados.
2. Aplicación debe:
 - a. Permitir que el usuario explore las características y variables que tiene el conjunto de datos (luego de las actividades de limpieza), especialmente con las que se crearon los modelos. Esto debe hacerlo en forma de dashboard con gráficas enlazadas.
 - b. Permitir al usuario que vea los resultados de cada algoritmo por separado o de todos.
 - c. Mostrar, de forma informativa, en gráficas interactivas el rendimiento de cada uno de los algoritmos con los datos de prueba. También debe permitir que el usuario decida si esconder o no esa información.
3. El diseño de la aplicación debe ser intuitivo y amigable con el usuario. Los colores deben ser cuidadosamente seleccionados para que cumpla con la teoría del color estudiada en clase.

Informe:

- Elabore un informe en el que reúna toda la información del proyecto. Debe tener las siguientes secciones:
 - Introducción
 - Se debe introducir el contenido del informe.
 - Se debe plantear el problema a resolver.
 - Objetivos
 - Tanto general como específicos.
 - Marco Teórico
 - Debe contener toda la teoría estudiada para elaborar el proyecto
 - En caso de procesamiento de imágenes:
 - Información sobre procesamiento de imágenes

- Información teórica sobre análisis de las imágenes desde el punto de vista médico o de la especialidad a la que hacen referencia.
 - Algoritmos de aprendizaje de máquinas, ya sea profundo o no que serían de utilidad para resolver el problema planteado.
 - En caso de procesamiento de lenguaje natural:
 - Definición de procesamiento de lenguaje natural,
 - Actividades de preprocesamiento
 - Algoritmos usados para procesamiento del lenguaje natural.
 - En caso de los sistemas de recomendación
 - Definición teórica de los sistemas de recomendación
 - Actividades de preprocesamiento general
 - Algoritmos usados para sistemas de recomendación.
- Metodología
 - Pasos que siguió el grupo para resolver el problema
 - Explicación de cómo seleccionó el grupo los conjuntos de entrenamiento y prueba.
 - Explicación de la selección de los algoritmos y las razones por las cuales los escogieron.
 - Explicación de selección de las herramientas utilizadas
 - Recursos de cómputo.
 - Lenguajes de programación, bibliotecas y/o paquetes utilizados.
- Resultados y Análisis de Resultados
 - Describe las características del conjunto de datos original
 - Describe las tareas de limpieza y preprocesamiento a las que tuvo que someter a los datos para lograr los resultados obtenidos.
 - Agrega el análisis exploratorio que se le practicó a los datos. Incluya gráficos que expliquen de manera visual los hallazgos encontrados en esta parte.
 - Explica el ajuste de los parámetros que hubo que hacerle a cada uno de los algoritmos para mejorar el rendimiento y la efectividad.
 - Compara los algoritmos de acuerdo con la efectividad, tiempos de procesamiento, errores, etc. Utiliza para esto gráficos explicativos, estáticos con colores adecuados
 - Describe la aplicación creada para probar los algoritmos, incluye capturas de pantalla y explica la selección de las tecnologías usadas para construirla.
- Conclusiones
 - Explica de qué forma se cumplieron cada uno de los objetivos planteados al inicio y discute sobre la efectividad de los algoritmos seleccionados para resolver el problema planteado.
- Bibliografía

- Se listan en esta sección los documentos consultados para elaborar el informe.
- Se referencian en el texto las referencias.

EVALUACIÓN

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

- **(15 puntos) Preprocesamiento de los datos de entrada:** Se deben realizar las tareas de preprocesamiento para preparar la entrada para que pueda ser analizada por la aplicación. Esto debe pasar inadvertido para el usuario.
- **(20 puntos). Presentación de resultados:** Se informa el resultado de la clasificación de la entrada al sistema usando modelos entrenados con varios algoritmos. Se puede comparar la eficiencia de cada uno de los modelos basado en la respuesta que da. Se puede seleccionar el modelo a utilizar o utilizarlos todos. La aplicación permite al usuario explorar de forma dinámica e intuitiva los datos.

NOTA: En la presentación de resultados, no se trata solo de mostrar el rendimiento o la precisión de los modelos entrenados. Es importante que las visualizaciones cuenten la historia detrás de los datos: cómo se comportan, qué patrones o relaciones se descubrieron y cómo se reflejan las predicciones o clasificaciones obtenidas. La comparación entre modelos puede incluir su eficiencia, pero también debe destacar los hallazgos más relevantes del análisis exploratorio y su impacto en la interpretación de los resultados. El objetivo es que las visualizaciones permitan al usuario comprender de forma clara, dinámica e intuitiva el proceso y las conclusiones del proyecto.

- **(15 puntos). Eficiencia de los modelos:** Se muestra la eficiencia de los modelos, usando gráficas interactivas. Se permite al usuario esconder esta información.
- **(40 puntos). Informe final:**
 - El informe tiene todas las secciones que se le solicitan en las instrucciones.
 - Las explicaciones son claras y coherentes.
 - La investigación está completa y bien estructurada.
- **(10 puntos). Referencias y Bibliografía:**
 - Construye la bibliografía consultada, siguiendo las normas APA.
 - Pone referencias consultadas en el texto.
 - La bibliografía referenciada está indexada por lo que es confiable

MATERIAL A ENTREGAR

- Archivo .pdf con el informe completo del proyecto.
- Scripts de R (.r o .rmd) o de Python que utilizaron para el proyecto.
- Link de Google Drive donde trabajó el grupo
- Link del repositorio usado para versionar el código.
- Presentación de Power Point a usar para presentar resultados.

FECHAS DE ENTREGA

- **AVANCES:** Debe entregar un documento con los avances reportados para el **10 de noviembre de 2025** con:
- **PRESENTACIÓN Y DOCUMENTO FINAL COMPLETO: 13 de noviembre de 2025**
- **NOTA:** Solo se calificará el Documento Final si está entregado el avance con todo lo que se pide.

REFERENCIAS

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