



CC219 - Aplicaciones de Data Science

Trabajo Parcial / Final

1. Objetivo

El presente documento define el trabajo Parcial/Final y la rúbrica que permite evaluar el logro del curso CC219 - Aplicaciones de Data Science.

2. Logro del curso

Competencia General: Pensamiento innovador

Nivel de logro: 2

Capacidad para generar propuestas sostenibles y creativas de solución a un problema, que implica la mejora o creación de un producto, servicio o proceso, impactando positivamente en un determinado contexto.

Competencia Especifica: ABET 5 - Trabajo Multidisciplinario

Nivel de logro: 1

Capacidad de trabajar en proyectos de equipo, que suelen ser multidisciplinarios. Esta base brinda a los estudiantes una amplia base de comprensión que les permite aplicar su conocimiento de los principios científicos a las soluciones prácticas e innovadoras de problemas existentes y futuros.

3. Enunciado

El Trabajo Parcial/Final consiste en aplicar las técnicas de Ciencia de Datos desarrolladas durante el presente ciclo de estudios, proponiendo un aporte de conocimiento, acorde con las competencias generales y específicas del curso.

Los alumnos elegirán un tema o caso de uso a desarrollar de acuerdo a su preferencia y pensamiento innovador.

4. Estructura del Informe

El informe debe contener como mínimo 6 páginas y máximo 15 páginas y escrito con la siguiente estructura:

Descripción del caso de uso. Redactar la descripción y fundamentación del problema y/o caso de uso del cual se propone encontrar algún conocimiento al ejecutar cada una de las tareas comprendidas durante las fases de un Proyecto de Ciencia de Datos (citar fuentes).



CC219 - Aplicaciones de Data Science

Trabajo Parcial / Final

El caso de uso debe contemplar como requerimiento el dar respuesta a mínimo tres preguntas que involucren una clasificación y/o predicción.

- **Descripción del conjunto de datos (dataset).** Redactar las características y origen de los datos recolectados motivo de análisis y para su posterior analítica.
- **Análisis exploratorio de los datos (EDA).** Los datos recolectados deberán ser semi o no estructurados. Se debe incluir la descripción de las tareas de carga, inspección, preprocesamiento y visualización de los datos.
- **Modelización.** Comprende el seleccionar uno más algoritmos para ser entrenados a partir de los datos preparados. Comprobar el rendimiento de los modelos creados y experimentar con ellos con datos de prueba para realizar clasificaciones / predicciones.
- **Publicación de los resultados.** Comunicar los resultados obtenidos a partir de los experimentos realizados con los modelos creados (uso de métricas y tablas comparativas).
- **Análisis de Explicabilidad.** Hacer el análisis del modelo indicando las variables más relevantes del modelo, así como la técnica de explicabilidad utilizada.
- **Conclusiones.** En un párrafo redactar las conclusiones del trabajo, especificando la(s) técnica(s) utilizadas, los resultados obtenidos (positivos o no), y de ser el caso, el trabajo a futuro.
- **Referencias bibliográficas**

5. Acerca del grupo de trabajo

El trabajo se deberá desarrollar en grupo de 3, según sea el caso.

6. Técnicas y temas aceptados

El conocimiento que se propone obtener resultará de la aplicación de una o varias de las siguientes técnicas de Ciencia de Datos desarrolladas en clase:

- **Minería de Textos**
 - Procesamiento de Lenguaje Natural (NLP) para hacer que grandes cantidades de datos no estructurados sean accesibles y útiles.
 - Clasificación de textos (uso / comparación de distintos algoritmos de aprendizaje automático, comparando sus rendimientos)



CC219 - Aplicaciones de Data Science

Trabajo Parcial / Final

Como ejemplos de temas a desarrollar dentro de la técnica de Minería de Textos pudieran ser: análisis de temas, análisis de sentimientos, detección de idioma y detección de intenciones.

- **Sistemas de Recomendación**

- Basados en Procesamiento de Lenguaje Natural o bien a partir de algoritmos de redes neuronales

- **Visión Computacional**

- Detección de objetos Tracking
- Facial Analysis

7. Lenguaje de programación

Las tareas de análisis y analítica deben estar desarrolladas en lenguaje Python salvo alguna restricción sustentada puede usar otro lenguaje

Los resultados obtenidos podrán ser visualizados a través de interfaz gráfica (GUI) simple y amigable, destinada a un usuario final (interesados del negocio o Stakeholders).

8. Exposición

- La exposición es parte de la evaluación del trabajo final que se realizará en la última sesión de la semana 15.
- Con Vestimenta formal.
- Una diapositiva máximo 15 páginas resaltando (problema y fundamento, estado de arte, propuesta, desarrollo del aplicativo si fuera el caso, resultados y discusión)
- 10 minutos de exposición por cada grupo.
- Se realizarán preguntas a los integrantes del grupo acerca del trabajo y deberán detallar el contenido.



CC219 - Aplicaciones de Data Science

Trabajo Parcial / Final

9. Evaluación del Trabajo Parcial/Final

El trabajo se ha dividido en 2 hitos:

PRIMER HITO (TRABAJO PARCIAL):

- **Escribir el informe** de acuerdo al siguiente contenido:
 - **Descripción del caso de uso.** Redactar la descripción y fundamentación del problema y/o caso de uso del cual se propone encontrar algún conocimiento al ejecutar cada una de las tareas comprendidas durante las fases de un Proyecto de Ciencia de Datos (Proyecto de Analítica)

El caso de uso debe contemplar como requerimiento el dar respuesta a mínimo dos preguntas que involucren una clasificación y/o predicción (citar fuentes).
 - **Descripción del conjunto de datos (dataset).** Redactar las características y origen de los datos recolectados motivo de análisis y para su posterior analítica.
 - **Análisis exploratorio de los datos (EDA).** Los datos recolectados deberán ser semiestructurados o no estructurados. Se debe incluir la descripción de las tareas de carga, inspección, preprocesamiento (o normalización en el caso de textos) y visualización de los datos.
 - **Propuesta de Modelización.** Comprende el seleccionar y proponer uno más algoritmos para ser entrenados a partir de los datos preparados a fin de resolver las preguntas de clasificación y/o predicción.
- **Entrega en Aula virtual:**
Crear un archivo con el nombre TP_XXX_YYY_ZZZ, donde los caracteres X, Y y Z corresponden al primer apellido de los integrantes del grupo.
 - Entregar en el Aula Virtual en Fecha: **26-09-2024 a las 23:59h**
 - Puntaje asignado: 0 a 20 puntos



CC219 - Aplicaciones de Data Science

Trabajo Parcial / Final

SEGUNDO HITO (TRABAJO FINAL):

- **Concluir el informe.**

El informe debe estar escrito de acuerdo al siguiente contenido:

- **Descripción del conjunto de datos (dataset).** Redactar las características y origen de los datos recolectados motivo de análisis y para su posterior analítica.
- **Análisis exploratorio de los datos (EDA).** Los datos recolectados deberán ser semi o no estructurados. Se debe incluir la descripción de las tareas de carga, inspección, preprocesamiento (o normalización en el caso de textos) y visualización de los datos.
- **Propuesta de Modelización.** Comprende el seleccionar y proponer uno más algoritmos para ser entrenados a partir de los datos preparados a fin de resolver las preguntas de clasificación y/o predicción.
- **Modelización.** Comprende el entrenamiento de uno más algoritmos propuestos a fin de obtener conocimiento a partir de los datos preparados. Comprobar el rendimiento de los modelos creados y experimentar con ellos con datos de prueba para realizar las clasificaciones / predicciones.
- **Publicación de los resultados.** Comunicar los resultados obtenidos a partir de los experimentos realizados con los modelos creados (uso de métricas y tablas comparativas).
- **Conclusiones.** En un párrafo redactar las conclusiones del trabajo, especificando la(s) técnica(s) utilizadas, los resultados obtenidos (positivos o no), y de ser el caso, el trabajo a futuro.
- **Referencias bibliográficas**

- **Entrega del trabajo final en el Aula virtual:**

Crear una carpeta con el nombre TF_XXX_YYY_ZZZ, donde los caracteres X, Y y Z corresponden al primer apellido de los integrantes del grupo. La estructura de la carpeta es el siguiente:

- ✓ Sub carpeta de “Código fuente”.
- ✓ Sub carpeta Dataset
- ✓ Link de video grabado de la exposición 12 minutos (cada uno 4 minutos) en un archivo de texto.
- ✓ TF_Informe_XXXXX_YYYYY.docx



CC219 - Aplicaciones de Data Science

Trabajo Parcial / Final

Empaquetar o comprimir la carpeta TF_XXX_YYY_ZZZ.zip entregar en Blackboard.

- Fecha: **Domingo de la semana 14 a las 23:59h**
- Rubrica de evaluación: **Rúbrica de trabajo final.**
- Puntaje asignado: 0 a 20 puntos.

Consideraciones adicionales

- Se deberá contemplar un repositorio en GitHub.com llamado **CC219-TP-TF-2024-2** conteniendo dos carpetas:
 - **data:** deberá contener el dataset original y el dataset final resultante (limpio o preparado para análisis).
 - **code:** deberá contener los programas en Python desarrollados.
- El archivo.Readme, dentro de GitHub, deberá contemplar:
 - Objetivo del trabajo
 - Nombre de los alumnos participantes
 - Breve descripción del dataset (se puede adjuntar el archivo PDF)
 - Conclusiones
 - Licencia
- En el documento entregable, se deberá **incluir el enlace a la cuenta de GitHub.com** desde donde se accede a la publicación de la evaluación.

Adicionalmente:

- Se evaluará el orden dentro de la organización del documento, así como la correcta redacción y gramática.
- Se valorarán las respuestas a preguntas que no hayan sido propuestas en la presente evaluación.
- Durante la semana 7 se realizará la exposición grupal.
- Cada grupo contará con 10-15 minutos para que cada integrante detalle cómo se obtuvieron las visualizaciones y conclusiones (preparar una presentación de no más de diez diapositivas).
- El orden de la exposición será mediante sorteo.
- El alumno que no se presente, perderá la calificación oral.
- La calificación podrá ser diferenciada por alumno si su participación / respuestas durante la exposición es insuficiente.