



CC221-Big Data

Trabajo Final

2025-01

1. Objetivo

El presente documento define el trabajo final y la rúbrica que permite evaluar el logro del curso CC221 – Big Data.

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 Específica: ABET 6 – Técnicas en computación

Nivel de logro: 1

La capacidad para comprender y brindar soporte para el uso, entrega y gestión de sistemas de información dentro de un entorno de sistemas de información

3. Enunciado

El trabajo final consiste en aplicar algoritmos de aprendizaje supervisado haciendo uso de la plataforma de Big Data H2O desarrolladas durante el presente ciclo de estudios, proponiendo un aporte de conocimiento, acorde con las competencias generales y específicas del curso.

4. Estructura del Informe

El informe debe contener 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 (citar fuentes).
- ✦ **Descripción del conjunto de datos obtenidos de kaggle.** Redactar las características y origen de los datos recolectados motivo de análisis y para su posterior analítica.
- ✦ **Análisis exploratorio de los datos (EDA).** Se debe incluir la descripción de las tareas de inspección, preprocesamiento, análisis univariado, bivariado y visualización de los datos.
- ✦ **Modelización.** Comprende la aplicación de los algoritmos de aprendizaje supervisado sobre la plataforma de Big Data llamada H2O y los compara.
- ✦ **Resultados.** Comunicar los principales resultados obtenidos (uso de métricas y tablas comparativas).
- ✦ **Conclusiones.** En un párrafo redactar las conclusiones del trabajo, especificando la técnica utilizada, los resultados obtenidos (positivos o no).
- ✦ **Recomendaciones.** Redactar los trabajos futuros.
- ✦ **Referencias bibliográficas**

5. Acerca del grupo de trabajo

El trabajo se deberá desarrollar en grupo de máximo 3 estudiantes.

6. Lenguaje de programación

Las tareas de análisis y analítica deben estar desarrolladas en el lenguaje Python sobre la plataforma de H2O.

7. Entrega en el aula virtual

Crear dos archivos con el siguiente nombre:

- **TF_XXX_YYY_ZZZ.docx** => hace referencia al informe
- **TF_XXX_YYY_ZZZ.ipynb** => hace referencia al notebook usado para la redacción del informe.

Los caracteres X, Y y Z corresponden al primer apellido de los integrantes del grupo.

- Exposición y entrega en el Aula Virtual en Fecha: **Domingo 13/07/2025 antes las 09:00h**
- Puntaje asignado: 0 a 20 puntos