



CC209-Data Mining Tools

Trabajo Parcial / Final

2024-02

1. Objetivo

El presente documento define el trabajo Parcial/Final y la rúbrica que permite evaluar el logro del curso CC209 – Data Mining Tools.

2. Logro del curso

Competencia General: Pensamiento

innovador Nivel de logro: 2

Capacidad para detectar necesidades y oportunidades para generar proyectos o propuestas innovadoras, viables y rentables. Planifica y toma decisiones eficientes orientadas al objetivo del proyecto.

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 Data Mining 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 escrito con la siguiente estructura:



CC209-Data Mining Tools

Trabajo Parcial / Final

2024-02

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 Data Mining (citar fuentes).



CC209-Data Mining Tools

Trabajo Parcial / Final

2024-02

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.
- **Ingeniería de Características.** Se deben utilizar y explicar técnicas de creación manual de características relacionadas al problema, mínimo 3 características por alumno con su respectiva justificación. Alguna técnica de selección de características o transformación de características para hacer la reducción de la dimensionalidad de los datos, se debe fundamentar y documentar la técnica o técnicas.
- **Prueba del proceso de selección de características:** Se debe usar algún algoritmo de aprendizaje básico para evaluar el proceso de ingeniería de características y explicar los resultados obtenidos.
- **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

Se debe desarrollar un problema que requiera aprendizaje supervisado, ya sea clasificación o regresión. El dataset debe ser una dataset tabular lo suficientemente grande. Mínimo 50 MBs.

7. Lenguaje de programación

Las tareas de análisis y analítica deben estar desarrolladas en lenguaje Python salvo



CC209-Data Mining Tools

Trabajo Parcial / Final

2024-02

alguna restricción sustentada puede usar otro lenguaje

8. Exposición

- La exposición es parte de la evaluación del trabajo parcial que se realizará en la última sesión de la semana 7.
- Una diapositiva máximo 15 páginas
- 10 minutos de exposición por cada grupo.
- Se realizarán preguntas a los integrantes del grupo acerca del trabajo y deberán detallar el contenido.

9. Evaluación del Trabajo Parcial

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: **25-09-2024 a las 23:59h**

Puntaje asignado: 0 a 20 puntos

10. Evaluación del Trabajo Final

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.
- **Propuesta de Modelización.** Comprende el seleccionar y proponer por lo menos 2 o 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 dos o 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.
- **Tuning de Hiperparametros.** Realizar una técnica avanzada de tuning de



CC209-Data Mining Tools

Trabajo Parcial / Final

2024-02

hiperparametros sobre los modelos propuestos.

- **Publicación del modelo.** Realizar una interfaz para utilizar los modelos.
- **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.

Entrega en Aula virtual:

Crear un archivo comprimido 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: **20-11-2024 a las 23:59h**

Puntaje asignado: 0 a 20 puntos

Referencias bibliográficas

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.