



Proyecto Ciencia de Datos

Resultados de Aprendizajes a Evaluar	<ul style="list-style-type: none">• Propone objetivos, preguntas e hipótesis a responder• Diseña e implementa proceso básico ETL• Desarrolla analítica exploratoria mediante técnicas estadísticas, datamining y machine learning.• Propone respuestas a las preguntas y evalúa hipótesis en base a los resultados obtenidos
Fecha Evaluación	27 de noviembre 2022 – Entrega de informe y Evidencia 29 de noviembre 2022 – Presentación Oral
Modalidad	Presencial
Forma de Entrega	<ul style="list-style-type: none">• Informe se entrega por plataforma eFinis• Presentación se entrega por plataforma eFinis• Evidencia (código y fuentes) se comparten por correo electrónico o enlace a: drojas@uft.cl, indicando el nombre de los integrantes del grupo.
Instrucciones Generales	<ul style="list-style-type: none">• La evaluación es grupal• Es responsabilidad del/la estudiante ser responsable de entregar el archivo correcto o cualquier acción que deba hacer para asegurar la entrega correcta de los entregables y respuestas solicitadas.• DEBE ENTREGAR TODOS LOS PROGRAMAS/DOCUMENTOS/ARCHIVOS EN UN SOLO ARCHIVO COMPRIMIDO .zip INDICANDO EL NOMBRE DE UN INTEGRANTE Y APELLIDO• Por cada día de atraso el descuento es de 1 punto (cada 24 horas desde fecha y hora del cierre de la evaluación)

Informe y Presentación

El informe es un documento redactado en tercera persona y la presentación es una exposición oral de todo el grupo el día de la evaluación. A continuación, la descripción y ponderación de cada apartado solicitado:

- **[1%] Tapa:** indicar, por lo menos, nombre del proyecto, integrantes y fecha.
- **[4%] Resumen:** Un resumen de media página de todo el contenido, principales propuestas, resultados y conclusiones.
- **[5%] Contexto:** Describir el contexto o dominio del proyecto.
 - Argumentar la importancia de realizar el proyecto
 - Argumentar sobre el impacto que podría tener su desarrollo.
 - Si existe un problema a resolver con el proyecto, deberá estar descrito en esta sección.
- **[10%] Propuesta:** La propuesta de su proyecto.
 - Proponer preguntas que permitan orientar el proyecto, preguntas que, para ser respondidas, permiten la generación de un objetivo y actividades a desarrollar.
 - Definir un objetivo general ¿qué se espera lograr con este proyecto?
 - Describir en forma general las actividades a desarrollar ¿qué se debe hacer para realizar el proyecto?
- **[15%] Preprocesamiento:** Describir la etapa de preprocesamiento
 - **Describir la adquisición de datos:** fuente, generación, disponibilidad, características, etc.
 - Describir la selección de datos: criterios de selección sobre la fuente original, criterios de limpieza y descarte.
 - **Describir la transformación de datos:** qué datos se transformaron, cambiaron tipo, cuantizaron, se crearon modelos representativos, se rellenaron o predijeron, etc.
 - **Describir la integración de datos:** si su proyecto utilizó más de una fuente, describir la forma en que se integran los datos, atributos compartidos, formas de relacionamiento, etc.
 - **Describir la carga de datos:** indicar como los datos son cargados en cada una de las distintas herramientas y técnicas utilizadas.
- **[25%] Análisis:** Analizar desde varias perspectivas. Se deben realizar todos los tipos de análisis, excepto el análisis de procesamiento de lenguaje natural que se descarta solo para los proyectos que no utilicen textos.
 - a. **Análisis Estadístico Descriptivo:** Describir mediante estadística descriptiva.
 - i. Describir por frecuencias (conteo), gráficos de barras, tablas de contingencia, etc.
 - ii. Describir estadígrafos generales como promedios, medias, mínimos, máximos, desviaciones estándar, etc.
 - b. **Análisis Exploratorio Estadísticos:** Describir el proceso de análisis factorial exploratorio sobre un subconjunto de las características de los datos cargados. Interpretar la confiabilidad de los resultados y los propios resultados según consideración de la técnica utilizada.

- c. **Análisis Exploratorio Procesamiento Lenguaje Natural:** Describir el procesamiento de lenguaje natural mediante la vectorización de textos y las características generales del modelo obtenido (vectores de documentos y palabras). Si es necesario, describir la reducción dimensional realizada. Interpretar los resultados obtenidos.
 - d. **Análisis Exploratorio Datamining:** Describir el proceso de análisis mediante reglas de asociación. Describir la forma en que se exploraron los distintos parámetros de la técnica e interpretar los resultados.
 - e. **Análisis Exploratorio Machine Learning:** Describir el proceso de análisis mediante clustering (utilizar al menos dos algoritmos distintos). Argumentar sobre la elección de los algoritmos y describir los parámetros utilizados. Interpretar los resultados según las técnicas utilizadas.
 - f. **Análisis Integrado:** Debe realizar **un análisis integrado** de, al menos, **dos tipos de análisis en conjunto**. Por ejemplo: análisis de procesamiento de lenguaje natural y clustering, análisis factorial exploratorio y reducción dimensional, etc.
- **[10%] Visualización:** Describir mediante gráficos y figuras los resultados de las distintas técnicas de análisis utilizadas o sus integraciones. Interpretar los resultados según su proyecto. Cada tipo de análisis puede generar sus propios gráficos.
 - **[10%] Conclusiones:** Describir los resultados que apoyan las respuestas a las preguntas, cumplen los objetivos y permiten evaluar la hipótesis de trabajo. Al concluir debe tener en cuenta las limitaciones de cada una de las técnicas e indicar explícitamente si cumplió con los objetivos de su proyecto.
 - **[20%] Presentación:** Describir los resultados que apoyan las respuestas a las preguntas, cumplen los objetivos y permiten evaluar la hipótesis de trabajo. Al concluir debe tener en cuenta las limitaciones de cada una de las técnicas e indicar explícitamente si cumplió con los objetivos de su proyecto.

Evidencia

Debe entregar evidencia de todos los análisis realizados. Esto incluye scripts, fuentes originales de datos (o enlace), resultados de procesamientos intermedios, gráficos generados, etc.

Estas evidencias deben ser enviadas como enlace o directamente (según sea la posibilidad) al correo drojas@uft.cl