



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

Minería de Datos IIC2433

Enunciado del Proyecto del Curso

Objetivos

El objetivo general del proyecto del curso es que puedan enfrentar un desafío real de Minería de Datos, profundizando los contenidos del curso y poniendo en práctica herramientas de análisis de datos.

Lograr un proyecto de alto estándar los pone en una excelente posición de cara al mercado, ya que podrán incluir en su portafolio personal la solución de un gran desafío. Para la entrega final del proyecto deberán tener un código bien estructurado y mantenible, que les aporte valor a sus portafolios.

Instrucciones

El trabajo pueden desarrollarlo en grupos de hasta dos personas. Los grupos deben quedar definidos para antes de la primera entrega. Cada proyecto tendrá un ayudante asignado al que pueden recurrir como guía para aclarar cualquier tipo de duda.

La temática y el alcance del proyecto serán propuestos por ustedes. Para dicha definición, deben tener en mente que el objetivo es que pongan en práctica (y/o extiendan) los conocimientos adquiridos en el curso en algún ámbito de su interés. Si ustedes lo desean, pueden utilizar modelos no estudiados dentro del curso. La propuesta que hagan debe considerar profundidad en al menos uno de las siguientes puntos: (1) pre-procesamiento y transformación, (2) modelamiento o (3) visualización.

Un ejemplo de proyecto que integre procesamiento y visualización podría ser, unir varias bases de datos que encuentre en la Web, preprocesarlas, visualizarlas y luego aplicar uno de los modelos vistos en clases sobre ellas. Un ejemplo de aporte en modelamiento sería hacer una revisión del estado del arte de un problema de minería de datos/machine learning, probar una variante de los modelos que se han utilizado hasta el momento y analizar los resultados.

Un proyecto enfocado en la visualización sería desarrollar un código que genere una visualización innovadora de los resultados típicos de un modelo de minería de datos para un contexto en particular,

por ejemplo, desarrollar una nueva visualización para apreciar mejor un conjunto de reglas de asociación.

El uso de paquetes de modelos queda restringido para aquellos no vistos en el curso y donde el foco no sea el modelo en sí. Esto debe ser consensuado con el ayudante asignado.

Dos posibles sitios para encontrar bases de datos útiles son: Kaggle y Google Dataset Search.

Este proyecto lleva adjunto el Código de Honor de la Escuela y Universidad. Cualquier infracción de ética, ya sea por plagio no citado o código compartido con otros grupos, significaría al menos un 1.0 en el curso y cita con el comité externo de ética, que analizará el caso en conjunto con la Dirección de Pregrado de la Escuela. Sea cauteloso en el uso de las referencias.

Evaluación

La evaluación del proyecto consiste en una entrega preliminar no evaluada, una entrega de avance y la entrega final:

- Entrega de Propuestas (obligatoria) - Fecha: 11 de Octubre, 2019. No entregar a tiempo implica una penalización de 1 punto en la escala de nota de la Entrega Final (es decir, nota máxima será un 6.0)
- Entrega 1 (30 %) - Fecha: 8 de noviembre, 2019.
- Entrega Final (70 %) - Fecha: 29 de noviembre, 2019.

Entrega de Propuestas

El objetivo de esta propuesta es verificar y modificar, de ser necesario, la envergadura del proyecto de acuerdo a las expectativas de trabajo. Deben entregar un informe que contenga:

- Nombres de los integrantes.
- Introducción: Contexto, propuesta y motivación.
- Marco Teórico: Deben presentar los conceptos importantes y la teoría en la que se basará el desarrollo y solución del proyecto. Qué modelos aplican en su solución, en qué parte, cómo y por qué (agregar referencias). No es necesario que entren en profundidad y total entendimiento, pero sí deben comprender el dominio del trabajo propuesto.
- Conclusión: Se espera que mencionen cuáles creen que podrían ser sus resultados, principales limitaciones y dificultades del proyecto.

Entrega 1

Deberán entregar un informe que explique lo que han realizado y cuáles han sido sus principales dificultades. Este informe es la base para el informe final. Se espera:

- Reformulación de la propuesta: Se debe adjuntar una versión actualizada de la propuesta inicial considerando el feedback de los ayudantes. En esta versión debe quedar explícito cual será el entregable final.
- Supuestos y decisiones.
- Revisión bibliográfica: Referenciar las fuentes de información que se hayan consultado, conjunto a su relevancia para el proyecto en cuestión.
- Código: Los resultados deben ser repetibles por lo que el código debe estar comentado al menos de manera general. En el informe tiene que quedar claro qué es lo que han implementado y qué no.
- Bases de Datos: Se debe entregar las bases de datos para poder repetir los experimentos. Indique link en Drive, Dropbox o coordine otro método de entrega.
- Dificultades: Explicite las dificultades técnicas y teóricas que se han presentado hasta ahora.
- Aprendizaje: : ¿Cuál es el mayor aprendizaje adquirido hasta ahora?

Se recomienda reunirse con los ayudantes designados antes de la Entrega Final.

Entrega Final

- La información asociada a la entrega final será publicada más adelante.

Forma de Entrega

Las entregas se realizarán por el Siding, en un archivo comprimido que contendrá el informe y código respectivo de la entrega. En el informe deben indicar la forma de acceso a las bases de datos utilizadas, como se mencionó en la descripción de la Entrega 1.

Política de Descuentos

En caso de atraso, se aplicará un descuento lineal de nota 7 a 1 en 24 horas en la Entrega 1 y Entrega Final.