

Proyecto de graduación Magíster en Ciencia de Datos (AFG Agosto 2024)

El Magíster en Ciencia de Datos UC Chile aporta conocimientos y técnicas avanzadas de la ciencia de datos para su aplicación en la solución a problemas en la academia, industria y el Estado, en todas las áreas del quehacer. Ciertamente, el área de ciencia de datos (data science) pasó en corto tiempo de ser un tópico de especialistas en estadística y computación a convertirse en tema de discusión masiva por sus profundas implicancias en la sociedad. Tales efectos son tan grandes que se ha llegado a hablar incluso de una nueva revolución similar a la industrial y a la del conocimiento.

El Magíster en Ciencia de Datos pertenece a la unidad académica del Instituto de Ingeniería Matemática y Computacional (IMC UC) de la Pontificia Universidad Católica de Chile (UC Chile), se dicta 100% online, a través de la plataforma Coursera y es interdisciplinario, participando también las facultades de Matemáticas e Ingeniería de la Pontificia Universidad Católica de Chile (UC Chile).

1. Descripción actividad final de grado del Magíster en Ciencia de Datos 🛭 🖳

Consiste en un Proyecto de Graduación desarrollado en forma grupal – de máximo 3 estudiantes –, cuyo término y evaluación final será individual.

El Proyecto de Graduación integra los conocimientos y competencias adquiridas a lo largo del programa, a través de un proyecto aplicado en una organización real – organismo público o privado, de la industria o de la academia – donde generarán soluciones a problemas profesionales o científicos concretos, con el acompañamiento de un profesor guía y ayudante experto.

La actividad final de graduación se integra de 3 actividades curriculares (cursos), desarrollados de manera bimestral:

- "MDS3050-Actividad de graduación I" (5 créditos, periodo 8)
- "MDS3060-Actividad de graduación II" (5 créditos, periodo 9)
- "MDS3070-Actividad de graduación III" (10 créditos, periodo 10)

Al término del octavo bimestre, cada grupo deberá presentar la propuesta del Proyecto desarrollada en "MDS3070-Actividad de graduación I" la que deberá incluir marco teórico, objetivos, metodología, resultados esperados y plan de trabajo, considerando los tiempos destinados para esta actividad curricular.

El escrito de su proyecto deberá ser entregado en un formato estándar que se proveerá para tales efectos, debe contener marco teórico, objetivos, metodología, resultados esperados, plan de trabajo, análisis de resultados, las conclusiones y discusión. Al término de estos cursos, el grupo deberá entregar su proyecto escrito para ser calificado.

El profesor guía de la Actividad de Graduación será designado por el Director del Programa y deberá cumplir los requisitos que para tal efecto establece el Reglamento del Programa.

La obtención del Grado de Magíster en Ciencia de Datos tendrá una calificación final que se obtendrá ponderando las notas parciales de acuerdo a la cantidad de creditaje de cada una.

2. Funcionamiento del proyecto

PARTICIPANTES:

- Equipo de trabajo integrado por mínimo 3 estudiantes, quienes desarrollan el proyecto específico en la
- · Profesor y ayudante experto del curso: contraparte desde el IMC y guía del equipo de trabajo a cargo del proyecto.

• Responsable de la organización, quien será su interlocutor y contraparte.

DURACIÓN DEL PROYECTO:

• Comprende el periodo de tiempo de acuerdo al bimestre académico.

EVALUACIONES:

- Durante este periodo habrá presentaciones de avances del proyecto evaluadas el equipo de estudiantes que integran el grupo.
- La evaluación final del proyecto incluye una presentación y un informe escrito, corresponde a una calificación para el equipo de trabajo, con evaluaciones propias del curso y evaluaciones desde la organización.
- Los equipos de trabajo también podrán realizar una evaluación hacia la organización, cuya finalidad es de retroalimentación para futuros trabajos colaborativos en proyectos de graduación del magíster.

OTRAS CONSIDERACIONES:

- Se recomienda una reunión semanal entre el equipos de trabajo y el equipo y la organización, en modalidad virtual.
- La metodología del equipo de trabajo dependerá de la naturaleza del proyecto y de la coordinación interna que se defina con la organización, siempre en conocimiento y consentimiento del profesor a cargo del curso y en concordancia con la carga académica correspondiente al curso.

3. Selección del proyecto	
 El IMC evalúa los antecedentes de la propuesta de proyecto completados en este formulario por parte de la organización (sección 8). El profesor a cargo del curso inicia una iteración con el fin de reunir más antecedentes o de realizar ajustes en la propuesta en caso de ser necesario. De acuerdo a los antecedentes e información recabada de la propuesta de proyecto, y se selecciona la propuesta que se alinea con la formación académica en la especialidad de ciencia de datos. El área de Gestión Docente del IMC se encarga de informar a la organización por la selección o no selección del proyecto. Plazo: 	ì
4. Confidencialidad proyecto	
Se proveerá de un acuerdo de confidencialidad para resguardar la información aportada por la organización.	
5. DESCRIPCIÓN DEL PROYECTO DE LA ORGANIZACIÓN	
1. Señalar información de quién presenta la propuesta de proyecto: * 👊	
Estudiante Magíster en Ciencia de Datos	
Profesor/a UC	
Otro	
2. En caso que el proyecto sea presentado por estudiante del Magíster en Ciencia de Datos. Indicasus datos (nombre completo y correo uc): * 🔍	ar
José Alexander Idárraga Cardona - jaidarragac@uc.cl	

3. Señalar tipo de vínculo con empresa/organización: * 🖫

Trabajador

4.	Señalar antigüedad (en meses) de vínculo entre estudiante y empresa/organización: * 🔲	
	228 meses, DBA(Administrator de Bases de Datos) Universidad Pontificia Bolivariana	
5.	3. Proyecto propuesto	□
5.	Temática del proyecto * □,	
	Análisis y predicción del rendimiento académico en la educación universitaria.	
6.	Nombre del proyecto * 🗔	
	Análisis y modelamiento del rendimiento académico aplicado a los estudiantes de pre grado de la U	
7.	Problema o necesidad que la empresa quiere resolver a través del proyecto * 🔲	
	a. En la formación académica, es de vital importancia para los educadores y la Universidad el poder entender mejor el desempeño académico de los estudiantes, así como también proveerles con un soporte y guía efectivos para completar los objetivos académicos de las asignaturas. Así mismo se busca identificar las características de los grupos sociales de los estudiantes en relación con su rendimiento académico.	
	b. Identificar a los estudiantes con riesgo académico en asignaturas específicas para predecir y tomar acciones preventivas que permitan ofrecer asesoramiento y apoyo adecuado.	
	c. Desarrollar un modelo para predecir el porcentaje de asistencia de los estudiantes y estimar su riesgo de deserción, con el objetivo de identificar estudiantes en riesgo y aplicar medidas preventivas para fomentar su permanencia en el programa académico.	
	d. Posibilidad de implementar modelos de deep-learnig debido al alto volumen de datos.	7
	Metodología general. Descripción de etapas, requerimientos de formación académica, de periodicidad de visitas a la empresa, de posible presentaciones del proyecto en la empresa, entrotros. *	е

Otro

a. Metodología general:

Como metodología inicial utilizaremos la CRISP-DM, la cual adaptaremos de acuerdo a las siguientes etapas. Así mismo emplearemos el desarrollo ágil para iterar rápidamente sobre los prototipos.

- b. Descripción de etapas:
- i. Solicitud de acceso a la información: Se realizarán las reuniones con las partes interesadas del proyecto para obtener el acceso a la información que será utilizada durante el desarrollo del proyecto.
- ii. Recolección de datos: Se implementarán las consultas necesarias para extraer los datos, asegurando la privacidad de los datos sensibles. Luego se recopilará los datos de encuestas de caracterización y notas académicas de los estudiantes. Asimismo se realizara la limpieza y preprocesamiento de datos, incluyendo la imputación de valores faltantes, normalización y codificación de variables categóricas.
- iii. Análisis exploratorio (EDA): Se procederá con la limpieza y el posterior análisis de los datos obtenidos para identificar patrones y relaciones entre variables. Igualmente se visualizarán los datos utilizando gráficos y diagramas para una mejor comprensión.
- iv. Modelamiento: Se implementará una técnica para el particionamiento de datos, ya sea usando validación cruzada o validación simple. Los modelos serán evaluados de acuerdo a las métricas: RMSE, MAE, R², accuracy según corresponda. Se empleará modelos de machine learning y deep learning para la representación de información de la caracterización. Así mismo, para el modelamiento ético se utilizarán modelos como LIME o SHAP.
- v. Análisis de resultados: Se analizarán los resultados obtenidos desde diferentes perspectivas, incluyendo los aspectos éticos involucrados.
- vi. Informe y Recomendaciones: Se elaborará un informe detallado con los hallazgos del análisis y las predicciones de los modelos, así como el impacto a nivel ético. Así mismo se propondrá recomendaciones basadas en los resultados para mejorar el rendimiento académico.
- c. Requerimientos de formación académica: Actualmente contamos en la UPB con las personas que nos pueden apoyar en la información que se requiera del sistema académico, las cuales tienen todo el conocimiento en el sistema de información, uno de los integrantes del proyecto es el DBA (José Alexander Idárraga Cardona) de la Universidad.
- d. Periodicidad de visitas a la empresa: Este apartado se coordinará previamente mediante reuniones con los encargados, estas reuniones se realizaran de forma virtual durante toda la ejecución del proyecto en función a la necesidad de recolectar datos adicionales.
- e. Presentaciones del proyecto en la empresa: Estas serán coordinadas con el área encargada para mostrar prototipos o insights significativos de acuerdo al avance del proyecto.

9.	Contenidos y herramientas para el desarrollo del proyecto. Elementos necesarios específicos que
	se utilizarán en el desarrollo del proyecto, como por ejemplo, uso de softwares en la empresa,
	conocimientos o competencias relacionados con el rubro de la empresa, etc. * 🗔

	a. Librerías de código abierto:
	i. scikit-learn ii. Pandas iii. Matplotlib iv. Pytorch v. SQL vi. Jupyter notebooks vii. nbdime
	b. Librerías propietarias:
	PLSQL
	c. Conocimientos o competencias relacionados con el rubro de la empresa:
	Aspectos académicos de los estudiantes de pre-grado de la UPB – Colombia
	d. Github para la gestión del proyecto
10. F	Periodo de duración del proyecto * 🗔 6 meses aproximadamente
11. (Comentarios (opcional) 🗔
	a. A continuación anexamos a los integrantes del proyecto con sus respectivos correos:
	i. José Alexander Idárraga Cardona jaidarragac@uc.cl ii. Darío Gonzalo Saavedra Contreras dgsaavedrac@uc.cl iii. Luis Antonio Aguilar Gutiérrez laaguilarg@uc.cl
	b. También anexamos un repositorio de Github:
	Repositorio https://github.com/LsAntonio/GradProjectPilot, en donde presentamos dos Notebooks con un análisis pre-eliminar de una muestra de los datos. El notebook de Análisis_encuestas contiene información relevante con respecto al modelamiento inicial del rendimiento académico

Gracias!

usando variables categóricas obtenidas de una encuesta. Mientras que el notebook Eda_present se

centra en estimar los promedios un solo curso de acuerdo a los componentes del curso.



Este contenido lo creó el propietario del formulario. Los datos que envíes se enviarán al propietario del formulario. Microsoft no es responsable de las prácticas de privacidad o seguridad de sus clientes, incluidas las que adopte el propietario de este formulario. Nunca des tu contraseña.

Microsoft Forms | Encuestas, cuestionarios y sondeos con tecnología de inteligencia artificial <u>Crear mi propio formulario</u> Privacidad y cookies | Términos de uso