

Programa de Asignatura



1 Identificación Asignatura

Nombre:	Ingeniería de Datos		Código:	IN1032
Carrera:	Ingeniería Civil Industrial	Unidad Académica:	Ciencias Naturales y Tecnología	
Ciclo Formativo:		Línea formativa:		
Semestre	VII	Tipo de actividad:	Obligatoria	
N° SCT:	6	Horas Cronológicas Semanales		
		Presenciales:	6	Trabajo Autónomo:
Pre-requisitos	IN1028 – Introducción a las TICA			

2 Propósito formativo

La asignatura de Ingeniería de Datos tiene como propósito que el estudiante desarrolle las habilidades necesarias para reconocer, gestionar y formular soluciones a partir de los datos generados en los diferentes sectores productivos. El reconocer el contexto bajo el cual es necesario almacenar, garantizar calidad y capitalizar el valor de los datos hacen que, en la actualidad, los ingenieros deban ser competentes en este ámbito.

El valor de los datos no es un concepto nuevo, ha sido ampliamente explorado en el contexto de las cadenas de valor de los datos, las cuales permiten tener la trazabilidad de los activos de datos, desde el momento que se adquieren, hasta el momento en que son explotados o generan un flujo de valor para las organizaciones. Los activos de datos son creados o utilizados en las organizaciones, en medio digital, papel u otros medios, y son considerados de gran valor por las empresas, por lo cual, son protegidos y custodiados.

Para cumplir con el propósito del curso, se realizará un acercamiento a las diferentes etapas en el ciclo de vida de los datos. En este sentido, la generación, obtención, administración, procesamiento y generación de soluciones a partir de los datos serán hitos claves para el desarrollo profesional en ingeniería. Este acercamiento permitirá llegar desde problemas y aplicaciones simples, en donde a partir de conocimientos iniciales de estadística y programación será posible formular soluciones alrededor del tema de la ingeniería de datos.

Esta asignatura aporta a la formación integral del estudiante desde diferentes perspectivas. Por un lado, los conocimientos sobre tratamiento, almacenamiento y programación entregados en esta asignatura le facilitarán gestionar proyectos relacionados, comprender la necesidad del contexto sobre el cual es necesaria la recolección de datos y garantizar en cierta manera, un aprovechamiento posterior de los mismos. Por otra parte, desde la ciencia de datos, dominar los fundamentos del aprendizaje supervisado y no supervisado para llevar a cabo una aplicación en datos reales durante el desarrollo del curso.

Esta asignatura profundiza en los contenidos vistos previamente en Herramientas Informáticas, Programación I, Programación II, Introducción a las TICA, Probabilidades y Estadísticas Aplicadas, en cuanto aprovecha de manera general conceptos de desarrollo, gestión y puesta en marcha de soluciones que requieren de un enfoque multidisciplinar para la generación de valor a partir de los datos. Si bien no es una línea de profundización, permitirá a los estudiantes acercarse con mayor familiaridad y estar en la capacidad de asumir una postura técnico – científica de cara a proyectos relacionados con la generación, administración, análisis y propuesta de soluciones a partir de los datos.

3 Contribución al perfil de egreso

Esta asignatura contribuye a los siguientes desempeños o resultados de aprendizaje globales declarados en el Perfil de Egreso de la carrera:

- Demuestra dominio de las ciencias básicas, de energía y de las ciencias de la ingeniería y la capacidad de realizar análisis de tipo económico, políticos y sociales.
- Concibe soluciones a los problemas que surgen en las organizaciones.
- Diseña, selecciona y adapta desarrollos tecnológicos y científicos propios de la ingeniería industrial a los desafíos de las organizaciones.

4 Resultados de aprendizaje específicos (verbo + objeto + condición + finalidad)

- RA1. Comprende los conceptos de aprendizaje, generalización, sobreajuste y robustez para aplicarlos satisfactoriamente a problemas supervisados y no supervisados.
- RA2. Reconoce los supuestos necesarios para poder aplicar un modelo específico en una situación concreta de negocio.
- RA3. Identifica los modelos de negocio, a través de los cuestionamientos adecuados, para formular soluciones con base en la aplicación de ingeniería de datos.
- RA4. Diferencia métodos y técnicas que están disponibles en los repositorios y librerías de desarrollo para modelar problemas de negocio.
- RA5. Aplica los fundamentos del aprendizaje supervisado y no supervisado para llevar a cabo una aplicación en datos reales durante el desarrollo del curso.

5 Unidades de Aprendizaje

1. Introducción a Ingeniería de datos.

- 1.1. Introducción a CRISP-DM
- 1.2. Concepto de Data Frame
- 1.3. Repositorio Git del curso
- 1.4. DataFrames de kaggle.com y pydataset.
- 1.5. Datos abiertos.
- 1.6. Lectura, limpieza y filtrado de datos
- 1.7. Escritura de datos
- 1.10. Visualización básica

2. Conceptos básicos de estadística y análisis exploratorio.

- 2.1. Estadísticos básicos y gráficas
- 2.2. Muestreo aleatorio y teorema central del límite
- 2.3. Contrastes de hipótesis
- 2.4. Correlación
- 2.5. Selección de atributos
- 2.6. Avance de proyecto bajo metodología CRIPS-DM

3. Análisis de Regresión.

- 3.1. Regresión lineal simple y regresión lineal múltiple
- 3.2. Regresión polinomial y regresión logística
- 3.3. Árboles de regresión

4. Clustering y Clasificación.

- 4.1. Concepto de distancia
- 4.2. Métodos de enlace
- 4.3. Clustering Jerárquico

- 4.4. K medoides y clustering espectral, K-means, K Nearest Neighbors kNN
- 4.5. Árboles de clasificación
- 4.6. Random Forest

5. Diseño y construcción de un Data Warehouse

- 5.1. Introducción y conceptos básicos
- 5.2. Arquitectura
- 5.3. Casos de estudio
- 5.4. NoSQL**

6. Big Data y Visualización de datos.

- 6.1. Almacenamiento de grandes cantidades de datos
- 6.2. Conceptos básicos en el ecosistema Hadoop
- 6.3. Configurar los productos y herramientas más usadas dentro del ecosistema Hadoop
- 6.4. Análisis y Visualización de Datos con Tableau
- 6.5. Manipulación de información para poder conectarnos a ella con Tableau

6 Recursos de Aprendizaje

Bibliografía:

- B1. Downey, A. (2012). Think Python. " O'Reilly Media, Inc.". Versión en español gratuita y digital: <http://sourceforge.net/projects/htmlcseifspa>
- B2. Han, J., Pei, J., & Kamber, M. (2011). **Data mining: concepts and techniques. Elsevier.**
- B3. Géron, A. (2019). **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.** O'Reilly Media.
- B4. McKinney, W. (2012). **Python for data analysis: Data wrangling with Pandas, NumPy, and IPython.** " O'Reilly Media, Inc."
- B5. Raschka, S., & Mirjalili, V. (2019). **Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2.** Packt Publishing Ltd.

Recursos materiales e infraestructura:

- B6. Laboratorio de computación.
- B7. Acceso a Ucampus.
- B8.** Computadores debidamente equipados para utilizar lenguajes de alto nivel (por ej.: Python, Tableau, otros).

7 Comportamiento y ética académica:

Se espera que los estudiantes actúen en sus diversas actividades académicas y estudiantiles en concordancia con los principios de comportamiento ético y honestidad académica propios de todo espacio universitario y que están estipulados en el *Reglamento de Estudiantes de la Universidad de Aysén*, especialmente aquéllos dispuestos en los artículos 23°, 24° y 26°.

Todo acto contrario a la honestidad académica realizado durante el desarrollo, presentación o entrega de una actividad académica del curso sujeta a evaluación, será sancionado con la suspensión inmediata de la actividad y con la aplicación de la nota mínima (1.0).

Planificación del curso

8 Responsables

Académico (s) Responsable (s) y equipo docente	Profesor: Johnny Valencia C		
Contacto	Correo: johnny.valencia@uaysen.cl		
Año	2023	Periodo Académico	Primer semestre
Horario clases	Cátedra: Lu 14.30 – 16.00 hrs. Mi 14.30 – 16.00 hrs. Sesión Prácticas: Mi 12.00 – 01.30 hrs.	Horario de atención estudiantes	Contactar previamente al profesor mediante email. Disponible Lu 16.00 – 18.00 Vi 14.30 – 16.00
Sala / Campus	Sala Virtual Ucampus – Google Sites – Sala A3		

9 Metodología de Trabajo:

En el curso se contemplan cuatro tipos de actividades docentes, las cuales se asocian a requerimientos de sala y al nivel de intervención del profesor/ayudante:

Actividad docente	Descripción	Intervención del profesor/ayudante	Requerimiento de sala
Exposición conceptual	El profesor introduce conceptos de programación preliminares y necesarios a otras actividades de índole práctica, de forma expositiva.	Alta	Sala de clases UCampus Online UCampus Offline
Programación expositiva	El profesor profundiza en la comprensión de elementos conceptuales a través de la exposición directa de la resolución de problemas de ingeniería de datos con ejemplos.	Alta	Sala de clases UCampus Online UCampus Offline
Programación tutorial	Funciona como la programación expositiva, pero el profesor realiza pausas para que los alumnos completen “pasos requeridos” antes de continuar. El objetivo es que todos los alumnos completen un paso definido por el profesor antes de continuar al siguiente.	Media	Laboratorio de computación Computador persona
Actividad práctica / Programación autónoma	Los estudiantes abordan y resuelven problemas de ingeniería de datos de forma autónoma, idealmente en grupos y sólo guiados por el profesor.	Baja	Laboratorio de computación Computador persona

En cualquier semana del semestre en curso se podría realizar una **evaluación menor** sobre las temáticas estudiadas a la fecha. Esta evaluación menor puede ser de los siguientes tipos:

- **Control:** evaluación individual o grupal, que se realiza en papel. Ocupará la primera parte de la clase (Quiz - 45 mins.).
- **Trabajo en clases:** Evaluación individual o grupal, que se realiza en computador. Ocupará la primera parte de la clase (45. mins.) o toda la clase.
- **Entrega tarea:** Evaluación grupal que se realiza en computador y fuera del horario de clases.

Como parte de la nota de presentación del curso, los estudiantes en parejas deberán hacer una **exposición de un proyecto de aplicación**, considerando su análisis, diseño e implementación de solución. Se entregará un instructivo específico para esta evaluación.

10 Evaluaciones:

Evaluación	Ponderaciones específicas	Ponderación nota presentación
Pruebas de cátedra	<ul style="list-style-type: none">• Prueba 1 (P1): 50%• Prueba 2 (P2): 50%	50%
Evaluaciones menores	El promedio simple entre las notas consideradas corresponde al 100%.	50%

Calificación final:

- Nota de presentación: 70%
- Examen Final: 30 % (Exposición de proyecto de aplicación)

Requisito de Aprobación

- Asistencia: 70%
- Nota Final: 4,0

11 Otros aspectos asociados al funcionamiento del curso:

- Toda la coordinación del curso (comunicaciones, actualización de notas, material, etc.) será realizada a través de UCampus. El estudiante deberá informar con tiempo suficiente si presenta dificultades de conexión para trasladar el requerimiento a la coordinación de programa.
- **Es deber del estudiante mantenerse informado de las noticias, avisos y material entregado por el profesor a través de estos medios, se sugiere instalar en su dispositivo móvil la aplicación de UCampus.**

12 Planificación de las actividades de enseñanza- aprendizaje y de evaluación

Semana	Resultado(s) de Aprendizaje (*)	Tema (Unidades de aprendizaje) y actividades	Recursos utilizados o lecturas (*)	Actividad(es)
1	RA1	1 Introducción a Ingeniería de datos. 1.1 Concepto de Data Frame 1.2 Repositorio Git del curso 1.3 DataFrames de kaggle.com y pydataset.	B1, ¡Error! No se encuentra el origen de la referencia.	
2	RA1	1 Introducción a Ingeniería de datos. 1.4 Datos abiertos. 1.5 Lectura, limpieza y filtrado de datos	B1, B2	
3	RA1	1 Introducción a Ingeniería de datos. 1.6 Escritura de datos 1.7 Visualización básica	B1, ¡Error! No se encuentra el origen de la referencia.	
4	RA1, RA3	2 Conceptos básicos de estadística y análisis exploratorio. 2.1 Estadísticos básicos y gráficas 2.2 Muestreo aleatorio y teorema central del límite	B1, ¡Error! No se encuentra el origen de la referencia.	
5	RA1, RA3, RA2	2 Conceptos básicos de estadística y análisis exploratorio. 2.3 Contrastes de hipótesis 2.4 Correlación 2.5 Selección de atributos	B1, ¡Error! No se encuentra el origen de la referencia.	
6	RA1, RA5, RA3 RA2	3 Análisis de Regresión. 3.1 Regresión lineal simple y regresión lineal múltiple 3.2 Regresión polinomial y regresión logística 3.3 Árboles de regresión	B1, ¡Error! No se encuentra el origen de la referencia.	Cátedra 1 (P1)
7	RA1, RA2	3 Análisis de Regresión. 3.2 Regresión polinomial y regresión logística 3.3 Árboles de regresión	B1, ¡Error! No se encuentra el origen de la referencia.	
8	RA1, RA5, RA3 RA2	4 Clustering y Clasificación. 4.1 Concepto de distancia 4.2 Métodos de enlace	B1, ¡Error! No se encuentra el origen de la referencia.	
9	RA1, RA5, RA3 RA2	4 Clustering y Clasificación. 4.3 Clustering Jerárquico 4.4 K medoides y clustering espectral, K-means, K Nearest Neighbors kNN	B1, ¡Error! No se encuentra el origen de la referencia.	
10	RA3, RA4, RA2	4 Clustering y Clasificación. 4.5 Árboles de clasificación 4.6 Random Forest	B1, ¡Error! No se encuentra el origen de la referencia., ¡Error! No se encuentra el origen de la referencia.	
11	RA3, RA4, RA2	5 Diseño y construcción de un Data Warehouse 5.1 Introducción y conceptos básicos 5.2 Arquitectura	B1, ¡Error! No se encuentra el origen de la referencia., ¡Error! No se encuentra el origen de la referencia.	Cátedra 2 (P2)

12	RA4, RA2	5 Diseño y construcción de un Data Warehouse 5.3 Casos de estudio 5.4 NoSQL		
13	RA3, RA4, RA2	6 Big Data y Visualización de datos. 6.1 Almacenamiento de grandes cantidades de datos 6.2 Conceptos básicos en el ecosistema Hadoop	B1, ¡Error! No se encuentra el origen de la referencia., ¡Error! No se encuentra el origen de la referencia.	
14	RA3, RA4, RA2	6 Big Data y Visualización de datos. 6.3 Configurar los productos y herramientas más usadas dentro del ecosistema Hadoop	B1, ¡Error! No se encuentra el origen de la referencia., ¡Error! No se encuentra el origen de la referencia.	
15	RA3, RA4, RA2	6 Big Data y Visualización de datos. 6.4 Análisis y Visualización de Datos con Tableau	B1, ¡Error! No se encuentra el origen de la referencia., ¡Error! No se encuentra el origen de la referencia.	
16	RA3, RA4, RA2	6 Big Data y Visualización de datos. 6.5 Manipulación de información para poder conectarnos a ella con Tableau	B1, ¡Error! No se encuentra el origen de la referencia., ¡Error! No se encuentra el origen de la referencia.	Exposición de un proyecto de aplicación
17				Exposición de un proyecto de aplicación
18		Periodo de exámenes		Exposición de un proyecto de aplicación

(*) Para referencias de Resultados de Aprendizaje y Recursos, consultar el programa del curso.