Universidad Católica del Norte

PROGRAMA DE ASIGNATURA

I. IDENTIFICACIÓN

Carrera o programa: Ingeniería en Tecnologías de Información				
Unidad responsable: Escuela de Ingeniería				
Nombre de la asignatura: Ciencia de datos				
Código: ECIN-00462				
Semestre en la malla[1]: 4				
Créditos SCT – Chile: 5				
Ciclo de Formación	Básico		Profesional	X
Tipo de Asignatura	Obligatoria	X	Electiva	
Clasificación de área de conocimiento[2]				
Área: Ingeniería y Tecnología		**Subárea:** Ingeniería Informática		
Requisitos:				
Prerrequisitos:		**Requisitos para:**		
● Estadística		● Ingeniería de Datos y Big Data		

II. ORGANIZACIÓN SEMESTRAL

Horas Dedicación Semanal (Cronológicas)	Docencia Directa	3	Trabajo Autónomo	5	Total	8	
Detalle Horas Directas	Cátedra	Ayudantía	Laboratorio	Taller	Terreno	Exp. Clínica	Supervisión
	1.5		1.5				

III. APORTE AL PERFIL DE EGRESO

La asignatura contribuye al dominio 1 del perfil de egreso, “Conocimiento científico y disciplinario”. Además, contribuye al dominio 2 “Habilidades y Actitudes Personales y Profesionales”. También contribuye al dominio 4 “Habilidades para la Práctica de la Ingeniería”. Al finalizar la asignatura las y los estudiantes serán capaces de aplicar técnicas de ciencia de datos y de sus disciplinas asociadas (estadística y aprendizaje automático) para modelar y resolver problemas aplicados en diversos contextos.

IV. COMPETENCIAS

La carrera declara las siguientes habilidades:
1.1. Aplicar conocimientos de matemáticas y ciencias naturales a la solución de problemas complejos de ingeniería.
1.2. Aplicar conocimientos de ciencias de la ingeniería a la solución de problemas complejos de ingeniería.
1.3. Aplicar conocimientos, métodos y herramientas con un enfoque sistémico principalmente en la ejecución de proyectos TI en el ámbito de las ciencias de la computación, infraestructura TI e ingeniería de software.
2.1. Identificar y resolver problemas con un razonamiento analítico.
2.2. Experimentar, investigar y descubrir conocimiento

[1] Este campo sólo se completa en caso de carreras con programas semestrales.

[2] Clasificación del curso de acuerdo a la OCDE

4.3. Concebir soluciones que involucren, por ejemplo, aplicaciones TI, infraestructura TI, toma de decisiones, gestión de datos y gestión de proyectos.
4.4. Diseñar soluciones que involucren, por ejemplo, aplicaciones TI, infraestructura TI, toma de decisiones, gestión de datos y gestión de proyectos.
4.5. Implementar soluciones que involucren, por ejemplo, aplicaciones TI, infraestructura TI, toma de decisiones, gestión de datos y gestión de proyectos.

V. RESULTADOS DE APRENDIZAJE

1. Identificar los conceptos fundamentales de Data Science y su relación con otras disciplinas.
2. Implementar métodos de obtención de datos en distintos dominios de aplicación
3. Distinguir las características de los datos a través de técnicas de visualización de datos
4. Diseñar modelos básicos para el análisis de datos en diversos contextos de aplicación
5. Aplicar modelos básicos de análisis de datos teniendo en consideración el dominio que está siendo analizado
6. Interpretar los resultados obtenidos mediante modelos de análisis de datos
7. Contribuir a la generación de acciones desde el rol que ejerce dentro del equipo, favoreciendo el aprendizaje colectivo.

VI. ÁREAS TEMÁTICAS

1. Introducción
1.1 Definición de Data Science
1.2 Metodologías
1.3 Relación con Estadística y Machine Learning
1.4 Aplicaciones

2. Obtención de datos
2.1 Data scraping
2.2 Repaso álgebra lineal
2.3 Grafos y procesamiento de redes
2.4 Procesamiento de lenguaje natural

3. Exploración y Visualización de Datos
3.1 Visualización de distribución de valores de una variable
3.2 Visualización de relaciones entre pares de variables
3.3 Visualización multidimensional de datos

4. Modelamiento estadístico y aprendizaje automático
4.1 Estadística descriptiva básica
4.2 Regresión lineal y clasificación lineal (SVM y regresión logística)
4.3 Modelamiento no lineal y regularización
4.4 Métricas para regresión y clasificación
4.5 Evaluación de modelos

5. Otras técnicas de modelamiento
5.1 Aprendizaje supervisado (árboles de decisión, K-NN y redes neuronales poco profun-

das)
5.2 Aprendizaje no supervisado (K-Means, PCA y reglas de asociación)

VII. ORIENTACIONES METODOLÓGICAS

1. La metodología a desarrollar en esta asignatura debe favorecer la interacción entre las y los estudiantes a través de trabajos prácticos colaborativos que permitan la solución a problemas específicos contextualizados a la asignatura.
• Se sugiere el uso de clases expositivas y participativas con un método combinado, es decir, clases expositivas intercaladas con trabajos en grupo de corta duración para responder preguntas.
• Se sugiere la utilización de la metodología activa de análisis de casos para desarrollar experiencias que permitan incorporar los elementos teórico-prácticos asociados a los resultados de aprendizaje de la asignatura.
2. Las experiencias de cátedra/laboratorio/taller deben ser realizadas por medio de la utilización de software moderno aplicable a la asignatura.
3. Se recomienda que las y los estudiantes realicen presentaciones periódicas sobre el trabajo realizado que incluya: contextualización, desarrollo y conclusiones.
4. Actividades prácticas recomendadas: cápsulas teóricas, reuniones de trabajo, taller de trabajo en equipo y liderazgo, presentaciones e informes escritos de avance en español, revisión del estado del arte asociado al problema, lluvia de ideas, análisis de alternativas y descripción detallada de la solución.

VIII. ORIENTACIONES Y CRITERIOS PARA EVALUACIÓN

1. Se recomienda la aplicación de una evaluación diagnóstica al inicio de la asignatura.
2. La asignatura podría contemplar dos instancias de evaluación de los resultados de aprendizaje: cátedra y taller/laboratorio.
• En el caso de existir, ambas debieran ser aprobadas por separado: el porcentaje de cada una de ellas deberá ser de 60% para cátedra y 40% para taller/laboratorio.
• En el caso que la asignatura tenga actividades de taller/laboratorio, éstas deben ser realizadas en grupos de estudiantes y se recomienda la elaboración por parte de los estudiantes de un informe sobre la actividad desarrollada.
3. Se evaluará el conocimiento conceptual y procedimental mediante la realización de al menos dos pruebas sumativas de carácter presencial.
• Se recomienda además la aplicación de una evaluación mediante la entrega de un trabajo desarrollado en las horas indirectas asociadas a la asignatura.
• Se recomienda que las y los estudiantes realicen una o más presentaciones de los trabajos realizados, la evaluación de la misma debe ser por medio de la aplicación de una rúbrica.
4. Se recomienda realizar evaluaciones de carácter formativo. Esto permite al docente introducir correcciones, añadir alternativas y reforzar los aspectos para ayudar al estudiantado en el logro de sus habilidades.
5. La asistencia y condiciones de aprobación de la asignatura debe ser acorde a la aplicación del Reglamento de Docencia de Pregrado.

IX. RECURSOS BIBLIOGRÁFICOS

Bibliografía Mínima:

- McKinney, W. (2022). Python for Data Analysis. Estados Unidos: O'Reilly Media.

- Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Suiza: O'Reilly Media.
- Anting, N. (2022). DATA SCIENCE with KNIME: Data Exploration, Machine Learning and Visualization using CODELESS Visual Programming. NASPSOFT Intelligence Academy.

Bibliografía Complementaria:

- James, G. M., Witten, D., Hastie, T., Tibshirani, R. (2017). An Introduction to Statistical Learning: With Applications in R. Estados Unidos: Springer.
- Witten, I. H., Frank, E., Hall, M. A., Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. Países Bajos: Elsevier Science.
- Provost, F., Fawcett, T. (2013). Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking. Estados Unidos: O'Reilly Media.
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
- Wilke, C. (2019). Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures. Taiwán: O'Reilly Media.