



PROGRAMA DE ASIGNATURA

Nombre del curso o actividad	MINERIA DE DATOS
Nombre del programa	Doctorado en Ciencias de la Ingeniería mención Automática
Descripción	<i>Código: 11340, Nivel: 2 o 3, Créditos SCT-Chile: 10, Tipo: Electiva, Plan de Estudios: Resolución N°1118 del 03/04/20.</i> Proporcionar a los(as) estudiantes los fundamentos teóricos y prácticos de la minería de datos, capacitándolos para descubrir patrones, tendencias y conocimientos significativos en grandes volúmenes de información mediante el uso de técnicas avanzadas de análisis de datos, aprendizaje automático y herramientas computacionales, con aplicación directa en la investigación científica y la toma de decisiones en ingeniería y automatización industrial.
Objetivos	Objetivo General: Desarrollar competencias avanzadas en minería de datos y descubrimiento de conocimiento, aplicando algoritmos de aprendizaje automático, técnicas estadísticas y herramientas computacionales para el análisis exploratorio, modelado predictivo y extracción de información significativa de grandes volúmenes de datos en contextos de ingeniería y automatización. Objetivos Específicos: • Comprender los fundamentos del proceso de descubrimiento de Conocimiento en Bases de Datos (KDD) y la metodología CRISP-DM. • Aplicar técnicas de preprocesamiento de datos incluyendo limpieza, transformación, reducción de dimensionalidad y selección de características. • Implementar algoritmos de aprendizaje supervisado para clasificación y regresión, evaluando su desempeño mediante métricas apropiadas. • Aplicar técnicas de aprendizaje no supervisado para agrupamiento (<i>clustering</i>) y detección de patrones en datos sin etiquetar. • Desarrollar proyectos de minería de datos utilizando herramientas como Python (<i>scikit-learn</i>, <i>pandas</i>), R, WEKA o RapidMiner, integrando el análisis de datos en soluciones de ingeniería.
Contenidos	Unidad 1: Fundamentos de Minería de Datos y Descubrimiento de Conocimiento • Introducción a la minería de datos y el proceso KDD (<i>Knowledge Discovery in Databases</i>). • Metodología CRISP-DM (<i>Cross-Industry Standard Process for Data Mining</i>) y DataPro. • Tipos de problemas en minería de datos: descriptivos, predictivos y prescriptivos. • Aplicaciones en ingeniería, automatización industrial y sistemas inteligentes. Unidad 2: Preprocesamiento y Preparación de Datos • Recolección y exploración inicial de datos. • Limpieza de datos: tratamiento de valores faltantes, <i>outliers</i> y datos inconsistentes. • Transformación de datos: normalización, estandarización y discretización. • Reducción de dimensionalidad: PCA (<i>Principal Component Analysis</i>), SVD (<i>Singular Value Decomposition</i>), NMF (<i>Non-Negative Matrix Factorization</i>). • Selección de características (<i>feature selection</i>) y extracción de características (<i>feature extraction</i>).



	Unidad 3: Técnicas de Aprendizaje Supervisado • Clasificación: árboles de decisión (ID3, C4.5, CART), k-NN (<i>k-Nearest Neighbors</i>), Naive Bayes. • Máquinas de Vectores de Soporte (SVM - <i>Support Vector Machines</i>). • Redes neuronales artificiales y perceptrón multicapa. • Métodos probabilísticos y modelos bayesianos. • Regresión: regresión lineal, regresión polinomial, regresión logística. • Evaluación de modelos: matrices de confusión, precisión, <i>recall</i>, F1-score, ROC-AUC, validación cruzada. Unidad 4: Técnicas de Aprendizaje No Supervisado • Algoritmos de agrupamiento (<i>clustering</i>): <i>k-means</i>, canopy, agrupamiento jerárquico (aglomerativo y divisivo). • DBSCAN (<i>Density-Based Spatial Clustering</i>) y métodos basados en densidad. • Mapas autoorganizativos (SOM - <i>Self-Organizing Maps</i>). • Detección de anomalías y <i>outliers</i>. • Métricas de evaluación: índice de silueta, coeficiente de Davies-Bouldin, índice de Dunn. Unidad 5: Herramientas y Aplicaciones Avanzadas en Minería de Datos • Herramientas de software: Python (<i>scikit-learn</i>, pandas, NumPy), R, WEKA, RapidMiner, Orange. • Integración con Big Data: <i>Apache Spark</i>, Hadoop, procesamiento distribuido. • Minería de textos y Procesamiento de Lenguaje Natural (NLP). • Minería de procesos (<i>process mining</i>) y análisis de flujos de trabajo. • Aplicaciones en ingeniería: mantenimiento predictivo, optimización de procesos, detección de fraudes, sistemas de recomendación. • Consideraciones éticas y de privacidad en el análisis de datos.
Modalidad de evaluación	Estrategias Metodológicas: • Clases expositivas: Presentación de fundamentos teóricos de minería de datos y descubrimiento de conocimiento mediante exposiciones magistrales interactivas. • Sesiones prácticas de laboratorio: Implementación de algoritmos de preprocesamiento, clasificación, regresión y <i>clustering</i> utilizando Python (<i>scikit-learn</i>, pandas) y herramientas especializadas (WEKA, RapidMiner). • Aprendizaje basado en proyectos: Desarrollo de proyectos completos de minería de datos sobre conjuntos de datos reales, siguiendo la metodología CRISP-DM. • Estudio de casos: Análisis de aplicaciones reales de minería de datos en ingeniería, manufactura, finanzas y sistemas inteligentes. • Revisión de literatura científica: Análisis crítico de artículos de investigación recientes sobre técnicas avanzadas de minería de datos y aprendizaje automático. • Trabajo autónomo: Estudio independiente, resolución de problemas, implementación de algoritmos y desarrollo de proyectos. Evaluación: • Tareas y trabajos prácticos: Implementación de algoritmos de preprocesamiento, clasificación, regresión y <i>clustering</i>. Evalúa correctitud técnica, calidad del código, análisis de resultados y capacidad de interpretación. • Proyecto integral de minería de datos: Proyecto completo aplicando la metodología CRISP-DM sobre un conjunto de datos real. Evalúa comprensión del problema de negocio,



	preprocesamiento de datos, selección y aplicación de técnicas apropiadas, evaluación de modelos, interpretación de resultados y documentación técnica. • Presentación oral y defensa: Exposición del proyecto, demostrando dominio conceptual, capacidad de síntesis, comunicación técnica efectiva y respuesta a preguntas críticas.
Resultados de aprendizajes esperados	Resultado de Aprendizaje General: Al finalizar el curso, el(la) estudiante será capaz de aplicar el proceso completo de descubrimiento de conocimiento en bases de datos, desde la comprensión del problema y preparación de datos hasta la implementación de algoritmos de aprendizaje automático supervisados y no supervisados, evaluando críticamente los resultados obtenidos y comunicando efectivamente los hallazgos mediante herramientas computacionales especializadas para la solución de problemas complejos en ingeniería y automatización. Resultados de Aprendizaje Específicos: • Aplicar la metodología CRISP-DM para estructurar proyectos de minería de datos, identificando objetivos de negocio, requisitos técnicos y criterios de éxito en contextos de ingeniería. • Ejecutar técnicas de preprocesamiento de datos incluyendo limpieza, transformación, normalización, reducción de dimensionalidad (PCA, SVD) y selección de características para preparar conjuntos de datos de calidad. • Implementar y evaluar algoritmos de clasificación y regresión (árboles de decisión, k-NN, SVM, redes neuronales, métodos bayesianos) seleccionando el modelo más apropiado según las características del problema y evaluando su desempeño con métricas estadísticas rigurosas. • Aplicar algoritmos de agrupamiento (<i>k-means</i>, <i>clustering</i> jerárquico, DBSCAN, SOM) para descubrir patrones y estructuras en datos no etiquetados, interpretando y validando los resultados obtenidos. • Desarrollar soluciones completas de minería de datos utilizando herramientas profesionales (Python/<i>scikit-learn</i>, R, WEKA), integrando preprocesamiento, modelado, evaluación y visualización de resultados en proyectos aplicados a problemas reales de ingeniería.
Bibliografía	Básica: • J. Han, M. Kamber, and J. Pei, <i>Data Mining: Concepts and Techniques</i>, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011. • I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, <i>Data Mining: Practical Machine Learning Tools and Techniques</i>, 4th ed. Burlington, MA, USA: Morgan Kaufmann, 2016. • G. James, D. Witten, T. Hastie, and R. Tibshirani, <i>An Introduction to Statistical Learning with Applications in R</i>, 2nd ed. New York, NY, USA: Springer, 2021. • Artículos científicos relacionados con el tema de tesis del(de la) estudiante. Recomendada: • D. T. Larose and C. D. Larose, <i>Discovering Knowledge in Data: An Introduction to Data Mining</i>, 2nd ed. Hoboken, NJ, USA: Wiley, 2014. • A. Géron, <i>Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow</i>, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2019. • W. McKinney, <i>Python for Data Analysis</i>, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.