



GOBIERNO DE LA CIUDAD DE BUENOS AIRES

Ministerio de Educación

AGENCIA DE HABILIDADES PARA EL FUTURO
Dirección de Educación Técnica Superior

Instituto de Formación Técnico Superior N° 18
Mansilla 3643 - C1425BBW - CABA

CARRERA TECNICATURA SUPERIOR en CIENCIAS DE DATOS E INTELIGENCIA ARTIFICIAL
Res. 3839-MEGC-2022

Ciclo Lectivo 2025

Materia: MODELIZADO DE MINERÍA DE DATOS

Curso: 2do Año **Comisión y Horario:** Martes y Miércoles de 1800Hs a 2215Hs

Profesor/a: Ing. Miguel Carlos Pita

Régimen: Primer/Segundo Cuatrimestre

Carga Horaria: 12Hs Cat./Sem.

1.- Fundamentación

La materia "Modelizado de Minería de Datos" en la Tecnicatura Superior en Ciencia de Datos e Inteligencia Artificial (CD e IA) consiste en preparar a los estudiantes para aplicar técnicas avanzadas de minería de datos en la resolución de problemas complejos del mundo real. Se cubren aspectos fundamentales y avanzados del modelado de datos, incluyendo preprocesamiento de datos, técnicas de clasificación y regresión, y análisis de diferentes casos de estudio acerca de aplicaciones de la minería de datos. Los estudiantes utilizarán herramientas de software específicas para implementar modelos y analizar resultados.

El propósito de esta materia es que los futuros Técnicos Superiores en CD e IA manipulen, exploren y preparen la fuente de información para posteriormente procesar y manejar los datos que surjan de ella. Esto implica que realicen modelos diferentes para detectar datos atípicos, efectuar predicciones de comportamiento de los datos y analizar los resultados.

Para la organización de la enseñanza se han organizado los contenidos en dos bloques: "**Procesamiento de Datos**" y "**Modelos de minería de datos**".

El bloque **Procesamiento de Datos** se enfoca en el proceso de extraer información útil y comprensible a partir de grandes volúmenes de datos con el objeto de predecir tendencias y comportamientos y/o descubrir modelos previamente desconocidos.

El bloque **Modelos de Minería de datos** consiste en generar modelos o patrones comprensibles de bases de datos mediante herramientas como árboles de decisión para la construcción de los modelos.

Esta organización de contenidos no implica que deban desarrollarse en ese orden. Los contenidos se abordan teniendo en cuenta los alcances de las capacidades a desarrollar y los problemas propios del campo profesional.



2.- Objetivos

El objetivo de esta materia es proporcionar a los estudiantes una comprensión profunda de las técnicas y métodos utilizados en la minería de datos, y cómo aplicar estos conocimientos en situaciones prácticas. Se espera que al finalizar el cursado del módulo los estudiantes sean capaces de:

- Preprocesar y limpiar conjuntos de datos.
- Comparar distintas técnicas de minería de datos e identificar la más apropiada de acuerdo al área de aplicación.
- Aplicar técnicas de minería de datos para clasificación, regresión y agrupamiento.
- Implementar y evaluar modelos de minería de datos utilizando herramientas de software.
- Manejar herramientas para la aplicación de técnicas de extracción de conocimiento en bases de datos.
- Detectar patrones y realizar la documentación técnica para apoyo a la toma de decisiones.
- Interpretar y comunicar los resultados de los análisis de datos.

3.- Contenidos

UNIDAD 1: INTRODUCCIÓN A LA MINERÍA DE DATOS

- Definición y objetivos de la minería de datos.
- Tipos de datos y fuentes de datos.
- Proceso de minería de datos: CRISP-DM.
- Herramientas y software para minería de datos:
 - Python 3: El lenguaje de programación principal.
 - Anaconda: Entorno de desarrollo.
 - Jupyter Notebook y Google Colab: Herramientas interactivas.
 - Power BI: Herramienta para la visualización de datos.
- Introducción a las librerías de Python para minería de datos:
 - Pandas, NumPy, Matplotlib, Seaborn, Plotly y Cufflinks.
 - Scikit-learn: Para el aprendizaje automático.
 - TensorFlow: Para el aprendizaje profundo.

UNIDAD 2: PREPROCESAMIENTO DE DATOS

- Limpieza y transformación de datos.
- Manejo de datos faltantes y ruidosos.
- Normalización y escalamiento de datos.
- Selección y extracción de características.
- Uso de librerías de Python:
 - Pandas y NumPy para manipulación y preprocesamiento.
 - Scikit-learn para normalización y selección de características.



GOBIERNO DE LA CIUDAD DE BUENOS AIRES

Ministerio de Educación

AGENCIA DE HABILIDADES PARA EL FUTURO
Dirección de Educación Técnica Superior

Instituto de Formación Técnico Superior N° 18
Mansilla 3643 - C1425BBW - CABA

UNIDAD 3: ALGORITMOS DE APRENDIZAJE SUPERVISADO

- **Conceptos fundamentales del aprendizaje supervisado.**
- **Técnicas de Clasificación:**
 - Árboles de decisión.
 - Máquinas de vectores de soporte (SVM).
 - Redes neuronales (con introducción a TensorFlow y Keras).
 - k-Nearest Neighbors (k-NN).
- **Técnicas de Regresión:**
 - Regresión lineal y no lineal.
 - Regresión logística.
- **Evaluación y validación de modelos supervisados.**
- **Implementación con Scikit-learn y TensorFlow.**

UNIDAD 4: ALGORITMOS DE APRENDIZAJE NO SUPERVISADO

- **Conceptos fundamentales del aprendizaje no supervisado.**
- **Técnicas de Clustering:**
 - K-Means.
 - Jerárquico.
- **Reducción de Dimensionalidad:**
 - Análisis de Componentes Principales (PCA).
- **Análisis de agrupamiento y segmentación de datos.**
- **Implementación con Scikit-learn y TensorFlow.**

UNIDAD 5: ALGORITMOS DE APRENDIZAJE POR REFORZAMIENTO

- **Conceptos fundamentales del aprendizaje por reforzamiento.**
- **Modelos de Aprendizaje por Reforzamiento:**
 - Q-Learning.
 - SARSA.
 - Deep Q-Networks (DQN) utilizando TensorFlow.
- **Exploración vs. Explotación en el aprendizaje por reforzamiento.**
- **Aplicaciones del aprendizaje por reforzamiento:**
 - Juegos, robótica, optimización.
- **Implementación básica con librerías como TensorFlow y OpenAI Gym.**

UNIDAD 6: APLICACIONES DE LA MINERÍA DE DATOS

- **Casos de estudio en diferentes industrias (salud, marketing, finanzas, etc.).**
- **Implementación de proyectos de minería de datos.**
- **Interpretación y comunicación de resultados utilizando Power BI.**



4.- Bibliografía:

BIBLIOGRAFÍA OBLIGATORIA	
Apuntes y archivos del Profesor	
Libro	Bishop, C. M. (2006). <i>Pattern Recognition and Machine Learning</i> . Springer.
Libro	Han, J., Pei, J., & Kamber, M. (2011). <i>Data Mining: Concepts and Techniques</i> (3rd ed.). Morgan Kaufmann.
Libro	Tan, P.-N., Steinbach, M., & Kumar, V. (2005). <i>Introduction to Data Mining</i> . Pearson.
Libro	Géron, A. (2019). <i>Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow</i> (2nd ed.). O'Reilly Media.

BIBLIOGRAFÍA COMPLEMENTARIA	
Libro	Murphy, K. P. (2012). <i>Machine Learning: A Probabilistic Perspective</i> . MIT Press.
Libro	Aggarwal, C. C. (2015). <i>Data Mining: The Textbook</i> . Springer.
Libro	Tan, P.-N., Steinbach, M., & Kumar, V. (2005). <i>Introduction to Data Mining</i> . Pearson.

5.- Modalidad de cursada:

La cursada de la materia consistirá en el desarrollo de las distintas Unidades didácticas. Las clases dictadas serán del tipo teóricas, teórico-prácticas y prácticas en las que el énfasis estará puesto en el trabajo del alumno y en el entendimiento de los distintos temas dados. Se hará hincapié principalmente en las habilidades que deben desarrollar los alumnos relacionadas con el modelizado de minería de datos como también en los hábitos de concentración, dedicación, utilización de la bibliografía y estudio de los alumnos. Se desarrollarán también distintos ejercicios, para que el alumno pueda completar su proceso de aprendizaje, mediante la ejecución de casos prácticos de aplicación para cada tema y un Trabajo Práctico Integrador en grupos, para que puedan poner en práctica de manera integral todos los conocimientos aprendidos.

Se utilizarán las siguientes modalidades de cursada:

- Modalidad sincrónica conectados vía remota: Material empleado, Notebook para el empleo de las diferentes herramientas de modelizado de minería de datos, Aula Virtual del Gobierno de la Ciudad de Buenos Aires y tableta gráfica. Se tomará asistencia al inicio y fin de la clase. Concurrencia Obligatoria.
- Modalidad Presencial: Se tomará asistencia al inicio y fin de la clase. Concurrencia Obligatoria.



- Requisitos de Regularidad: Asistencia del 75% de las clases, condición de aprobado en el Trabajo Práctico Integrador y obtener más de 4 (cuatro) puntos en el examen parcial o recuperatorio (para los alumnos que hayan desaprobado el examen parcial). La nota de cursada para aquellos alumnos que realicen y aprueben el examen recuperatorio será de 4 (cuatro) puntos, cualquiera sea la calificación que hayan obtenido en el recuperatorio.

6.- Condiciones de promoción

El plan de estudios de la Tecnicatura Superior en Ciencia de Datos e Inteligencia Artificial (CD e IA) no contempla condiciones de promoción.

7.- Evaluación

Conjuntamente al desarrollo de las clases, se impartirá un Trabajo Práctico Integrador, el cual será realizado de forma grupal. Este trabajo consistirá en distintas tareas que se van a ir presentando a lo largo de las clases, conforme los temas de la materia vayan avanzando. La evaluación de dichas actividades por parte del docente será continua y, la aprobación de la materia está supeditada a la aprobación de dicho trabajo práctico.

Además, se efectuará un examen parcial para comprobar el entendimiento por parte de los alumnos de los distintos temas desarrollados. Para los alumnos que desaprueben el examen parcial se tomará un recuperatorio del mismo.

Finalmente, se desarrollará un examen final, el mismo será individual e integrará todos los contenidos teóricos y prácticos de la cursada. Dicho examen final será de carácter teórico-práctico y escrito. En aquellos casos que puedan presentar duda respecto del nivel de conocimiento del alumno, la mesa examinadora podrá contemplar sumar una instancia oral al examen escrito realizado.

Para acceder al examen final, tanto el Trabajo Integrador como el examen parcial (o recuperatorio para los alumnos que hayan desaprobado el examen parcial) deben estar aprobados, así también como contar con la asistencia al 75% de las clases como mínimo.