



*Ministerio de Capital Humano  
Universidad Tecnológica Nacional  
Rectorado*

## **APRUEBA CURSO DE POSGRADO**

### **DOCTORADO EN INGENIERÍA, MENCIÓN SISTEMAS DE INFORMACIÓN**

Buenos Aires, 11 de diciembre de 2025

VISTO la Resolución N° 2034/25 del Consejo Directivo de la Facultad Regional Córdoba, a través de la cual se solicita la aprobación y autorización de implementación del Curso de Posgrado “Machine Learning, AutoML e Interpretable Machine Learning aplicados a Minería de Datos” para el Doctorado en Ingeniería, mención Sistemas de Información, y

#### **CONSIDERANDO:**

Que el Consejo Superior autorizó mediante Resolución N° 275/13, el dictado del Doctorado en Ingeniería, mención Sistemas de Información en la Facultad Regional Córdoba.

Que el curso propuesto pretende formar investigadores/as capaces de diseñar, implementar y evaluar sistemas de Machine Learning eficientes y transparentes, integrando tres dominios complementarios: Machine Learning como fundamento, Automated Machine Learning (AutoML) como democratización, e Interpretable Machine Learning (IML) como imperativo ético.

Que la Facultad Regional Córdoba cuenta con un plantel docente de elevado nivel académico y profesional, además de una prolongada y amplia experiencia en el dictado de cursos y seminarios vinculados al propuesto.

Que la Comisión de Posgrado de la Universidad ha analizado los antecedentes que acompañan la solicitud y avala la presentación, y la Comisión de Ciencia, Tecnología y Posgrado recomienda su aprobación.

Que el dictado de la medida se efectúa en uso de las atribuciones otorgadas por el Estatuto Universitario.



*Ministerio de Capital Humano  
Universidad Tecnológica Nacional  
Rectorado*

Por ello,

EL CONSEJO SUPERIOR DE LA UNIVERSIDAD TECNOLÓGICA NACIONAL

ORDENA:

ARTÍCULO 1º.- Aprobar el currículum del Curso de Posgrado “Machine Learning, AutoML e Interpretable Machine Learning aplicados a Minería de Datos” para el Doctorado en Ingeniería, mención Sistemas de Información que figura en el Anexo I y es parte integrante de la presente Ordenanza, en el marco de lo establecido por el Reglamento de la Educación de Posgrado de la Universidad, Ordenanza N° 1924, y en un todo de acuerdo con la Resolución C.S. N° 275/13.

ARTÍCULO 2º.- Autorizar el dictado del mencionado Curso en la Facultad Regional Córdoba, con el cuerpo docente que figura en el Anexo II y es parte integrante de la presente Ordenanza.

ARTÍCULO 3º.- Establecer que la propuesta mencionada en el Artículo precedente quedará supeditada al cronograma de dictado de las correspondientes actividades académicas de la Facultad Regional.

ARTÍCULO 4º.- Regístrese. Comuníquese y archívese.

ORDENANZA N° 2200

UTN
p.f.d.
l.p.
m.m.m.



Ministerio de Capital Humano  
Universidad Tecnológica Nacional  
Rectorado

**ORDENANZA N° 2200**

**ANEXO I**

**CURSO DE POSGRADO**  
**DOCTORADO EN INGENIERÍA EN SISTEMAS DE INFORMACIÓN**  
**“MACHINE LEARNING, AUTOML E INTERPRETABLE MACHINE LEARNING**  
**APLICADOS A MINERÍA DE DATOS”**

**1. FUNDAMENTACIÓN**

El curso forma investigadores capaces de diseñar, implementar y evaluar sistemas de Machine Learning eficientes y transparentes, integrando tres dominios complementarios:

- Machine Learning como fundamento: Principios estadísticos y algorítmicos del aprendizaje supervisado y no supervisado, base teórica para comprender cómo los sistemas aprenden de los datos.
- Automated Machine Learning (AutoML) como democratización: Automatización de procesos complejos (preprocesamiento, selección de algoritmos, optimización de hiperparámetros), permitiendo que profesionales con conocimientos básicos resuelvan problemas sofisticados de minería de datos, respondiendo a la escasez global de científicos de datos.
- Interpretable Machine Learning (IML) como imperativo ético: Garantiza transparencia y explicabilidad, abordando exigencias regulatorias (GDPR), requisitos éticos de fairness, y necesidades de confianza de stakeholders no técnicos.

**2. JUSTIFICACIÓN**

El Machine Learning se ha consolidado como tecnología transformadora impactando todos los



*Ministerio de Capital Humano  
Universidad Tecnológica Nacional  
Rectorado*

sectores. El AI Index Report de Stanford documenta inversión global sin precedentes en 2015-2020, generando demanda crítica de profesionales especializados.

Problemática de escasez de talento: El científico de datos debe dominar múltiples roles: ingeniero de datos, analista estadístico, ingeniero de ML, especialista en visualización y comunicador. Esta multiplicidad genera escasez severa, particularmente en sectores con menor capacidad económica (educación pública, ONGs, PyMEs, gobiernos locales). Las empresas tecnológicas concentran el talento mediante compensaciones que otros sectores no pueden igualar, creando una paradoja: los sectores que más se beneficiarían del análisis de datos son los que menos acceso tienen a recursos especializados.

AutoML como solución democratizadora: Frameworks como Auto-Sklearn, H2O AutoML, PyCaret y AutoGluon alcanzan resultados competitivos con soluciones manuales, en fracción del tiempo y sin conocimiento especializado profundo.

Imperativo de interpretabilidad: La automatización introduce opacidad. Los modelos operan como "cajas negras" con decisiones estadísticamente precisas pero incomprensibles, generando problemas técnicos (dificultad detectar sesgos), éticos y legales (decisiones sin justificación afectan derechos fundamentales), y de confianza. El Interpretable Machine Learning aborda esta problemática mediante técnicas como SHAP y LIME.

Integración AutoML + IML: Representa un paradigma emergente esencial: automatizar eficiencia sin sacrificar transparencia. Sin embargo, esta integración no es estándar en práctica profesional ni formación académica.

Alineación con políticas globales: Consenso de Beijing (UNESCO, 2019), recomendaciones IESALC, y Ethics Guidelines for Trustworthy AI (Comisión Europea, 2019) establecen requisitos de transparencia, explicabilidad y accountability.

Contribución al estado del arte: Abundantes trabajos aplican ML en educación, algunos aplican AutoML, escasos aplican IML, pero hay ausencia de propuestas integrales



*Ministerio de Capital Humano  
Universidad Tecnológica Nacional  
Rectorado*

combinando AutoML + IML sistemáticamente. El Modelo MAMDE (distinguido ACOFI 2025) representa contribución original.

La convergencia de abundancia de datos, disponibilidad de herramientas, exigencias regulatorias y demandas sociales de equidad algorítmica crea un momento donde la formación doctoral debe evolucionar. Este curso forma investigadores capaces de automatizar eficiencia manteniendo transparencia, con énfasis en aplicaciones de alto impacto social.

### **3. OBJETIVOS**

Que al finalizar el curso el/la estudiante sea capaz de:

- Comprender fundamentos teóricos y prácticos de Machine Learning supervisado y no supervisado.
- Dominar el flujo de trabajo completo de un proyecto de ML (pipeline/MLWF).
- Aplicar técnicas de AutoML para automatizar preprocesamiento, selección de algoritmos y optimización de hiperparámetros.
- Implementar métodos de IML para garantizar explicabilidad y transparencia.
- Diseñar soluciones integrales articulando eficiencia (AutoML) con transparencia (IML).
- Evaluar críticamente sesgos algorítmicos, aspectos éticos y cumplimiento regulatorio.
- Aplicar conocimientos a problemas reales en diversos dominios, con especialización en educación.
- Desarrollar competencias para comunicar resultados técnicos a audiencias no especializadas.

### **4. CONTENIDOS MÍNIMOS**

#### **Unidad 1: Fundamentos de Machine Learning y Minería de Datos**

1.1. Introducción a Machine Learning: Aprendizaje supervisado, no supervisado, por refuerzo.



*Ministerio de Capital Humano  
Universidad Tecnológica Nacional  
Rectorado*

Tipos de problemas: clasificación, regresión, clustering, reducción de dimensionalidad. Overfitting, underfitting, trade-off sesgo-varianza.

1.2. Pipeline de Machine Learning: Etapas: comprensión del problema, adquisición de datos, preprocesamiento, modelado, evaluación, despliegue. Metodologías: KDD, CRISP-DM, SEMMA. Rol del científico de datos.

1.3. Preprocesamiento de Datos: Data munging: limpieza, normalización, transformación. Tratamiento de valores faltantes y outliers. Feature engineering: creación, selección, extracción. Balanceo de clases: SMOTE, undersampling, oversampling. Escalado: StandardScaler, MinMaxScaler.

1.4. Algoritmos de ML Supervisado: Regresión: lineal, logística, Ridge, Lasso. Árboles de decisión y Random Forests. Support Vector Machines. K-Nearest Neighbors. Gradient Boosting, Redes neuronales básicas.

1.5. Algoritmos de ML No Supervisado: Clustering: K-means, DBSCAN, clustering jerárquico. Reducción de dimensionalidad: PCA, Reglas de asociación.

1.6. Validación y Evaluación: Train/test split, validación cruzada: k-fold, stratified k-fold. Métricas clasificación: accuracy, precision, recall, F1-score, ROC-AUC. Métricas regresión: MSE, RMSE, MAE, R<sup>2</sup>. Matriz de confusión, curvas de aprendizaje.

## **Unidad 2: Automated Machine Learning (AutoML)**

2.1. Motivación y Componentes: Problema de escasez de científicos de datos. Componentes: preprocesamiento automático, selección de algoritmos, optimización de hiperparámetros. Ventajas y limitaciones.

2.2. Preprocesamiento Automático: Feature preprocessing automático. Codificación de variables categóricas: one-hot encoding, target encoding, feature hashing. Manejo automático de missing values. Selección automática de transformaciones.

2.3. Selección Automatizada de Algoritmos: Meta-aprendizaje: aprender de experiencias



*Ministerio de Capital Humano  
Universidad Tecnológica Nacional  
Rectorado*

previas. Neural Architecture Search (NAS) para deep learning. Estrategias de exploración del espacio de algoritmos.

2.4. Optimización de Hiperparámetros: Grid Search, Random Search. Optimización Bayesiana: procesos gaussianos, TPE. Técnicas evolutivas: algoritmos genéticos. Hyperband y BOHB. Definición del espacio de búsqueda y función objetivo.

2.5. Ensemble Automático: Stacking, blending, voting. Construcción automática de ensembles.

2.6. Frameworks y Herramientas: Open Source: Auto-Sklearn, TPOT, H2O AutoML, PyCaret, AutoGluon, Auto-Keras. Comerciales: Google Cloud AutoML, Azure Automated ML, DataRobot, H2O Driverless AI.

### **Unidad 3: Interpretable Machine Learning (IML)**

3.1. El Problema de la Caja Negra: Trade-off entre precisión e interpretabilidad. Necesidad de transparencia: stakeholders, aspectos legales, confianza. Riesgos: sesgos, discriminación, decisiones injustas.

3.2. Conceptos Fundamentales: Interpretabilidad vs. explicabilidad. Interpretación global vs. local. Model-specific vs. model-agnostic methods.

3.3. Taxonomía de Métodos: Por alcance: global, local. Por etapa: pre-modelo, intrínseco, post-hoc. Por modelo: específico, agnóstico.

3.4. Métodos Pre-Modelo: Análisis Exploratorio (EDA): visualizaciones, análisis estadístico. Herramientas: pandas-profiling, sweetviz, dataprep. Selección de características: filtro, wrapper, embedded.

3.5. Modelos Intrínsecamente Interpretables: Modelos lineales: regresión lineal y logística, GLM, GAM. Modelos basados en árboles: árboles de decisión, feature importance. Modelos basados en reglas. Visualizaciones: curvas de validación, matrices de confusión, ROC curves.

3.6. Métodos Post-Hoc (Model-Agnostic): Interpretación global: Partial Dependence Plots



*Ministerio de Capital Humano*  
*Universidad Tecnológica Nacional*  
*Rectorado*

(PDP), ICE, Feature Interaction plots, Permutation Feature Importance, Global Surrogate Models. Interpretación local: LIME, SHAP.

3.7. Herramientas: SHAP library, LIME library, InterpretML (Microsoft), ELI5, Alibi (Seldon), What-If Tool (Google), PipelineProfiler, Yellowbrick.

#### **Unidad 4: Aplicación a Minería de Datos Educativa**

4.1. Contexto de Big Data Educativo: Fuentes de datos: LMS, sistemas de gestión estudiantil, evaluaciones en línea. Problemáticas: rendimiento académico, deserción, procrastinación, calidad educativa. Sistemas tradicionales vs. basados en computadora. Learning Analytics y Educational Data Mining.

4.2. El Modelo MAMDE (Modelo de Automatización para Minería de Datos Educativa):

Etapas 1: Análisis del Entorno, Definición del Problema y Preparación de Datos - Análisis del entorno educativo y definición de problemas. Configuración de objetivos. ETL automatizado para integración de datos heterogéneos. EDA aplicado a datos educativos. Preprocesamiento específico del dominio. Construcción de datasets educativos.

Etapas 2: Modelado de Datos y Optimización de Hiperparámetros - Feature engineering para datos académicos. Selección de características relevantes. Aplicación de AutoML a problemas educativos. Muestreo y validación cruzada.

Etapas 3: Interpretabilidad, Explicabilidad y Conclusiones - Interpretación de modelos para stakeholders educativos (directivos, docentes, estudiantes). Explicabilidad de predicciones individuales. Detección de sesgos en sistemas educativos automatizados. Conclusiones accionables para instituciones.

4.3. Casos de Estudio: Predicción de deserción estudiantil. Identificación de estudiantes en riesgo académico. Sistemas de recomendación académica. Evaluación de calidad educativa. Análisis de rendimiento académico.

4.4. Aspectos Éticos y Regulatorios: Recomendaciones de UNESCO (Consenso de Beijing) e



*Ministerio de Capital Humano  
Universidad Tecnológica Nacional  
Rectorado*

IESALC. Privacidad de datos estudiantiles. Fairness y equidad en sistemas educativos automatizados. Transparencia y accountability.

4.5. Trabajo Práctico Integrador: Aplicación completa del modelo MAMDE con datos reales. Pipeline end-to-end: desde datos crudos hasta conclusiones interpretables. Presentación de resultados a audiencia no técnica.

## **5. DURACIÓN**

El curso tendrá una duración de SESENTA (60) horas.

## **6. METODOLOGÍA**

Metodología mixta teórico-práctica con aprendizaje experiencial, la cual consistirá en 40% sesiones teóricas (conceptos fundamentales, marcos teóricos, algoritmos; clases expositivas con soporte audiovisual; discusión de fundamentos matemáticos y análisis de papers); 40% prácticas de laboratorio (implementación hands-on con Python, Jupyter Notebooks; ejercicios progresivos con datasets reales; Frameworks, Scikit-learn, Auto-Sklearn, TPOT, H2O AutoML, PyCaret, SHAP, LIME, PipelineProfiler, Yellowbrick; Debugging colaborativo) y 20% resolución de caso real propuesto por los y las estudiantes.

## **7. EVALUACIÓN Y APROBACIÓN**

La evaluación se realizará por medio de la entrega de un Proyecto Integrador final, en el cual se deberá resolver un problema real de minería de datos, preferentemente del campo profesional o interés del/la estudiante. El requisito de aprobación será la obtención de una calificación mínima 7/10 y la presentación oral del Proyecto.



*Ministerio de Capital Humano  
Universidad Tecnológica Nacional  
Rectorado*

**ORDENANZA N° 2200**

**ANEXO II**

**CURSO DE POSGRADO  
DOCTORADO EN INGENIERÍA EN SISTEMAS DE INFORMACIÓN  
“MACHINE LEARNING, AUTOML E INTERPRETABLE MACHINE LEARNING  
APLICADOS A MINERÍA DE DATOS”  
FACULTAD REGIONAL CÓRDOBA**

***Cuerpo Docente***

- Dr. Germán Antonio MONTEJANO (DNI 16.772.683)
- Dra. Ana Gabriela GARIS (DNI 25.700.407)
- Dr. Gabriel Antonio NOVILLO RANGONE (DNI 23.066.108)

-----