



*Ministerio de Capital Humana
Universidad Tecnológica Nacional
Facultad Regional Córdoba*

CÓRDOBA, 22 de octubre de 2025

VISTO, el expediente por el cual el Secretario de Posgrado solicita el aval académico de este Consejo para la aprobación del curso de posgrado "Machine Learning, AutoML e Interpretable Machine Learning aplicados a Minería de Datos"; y

CONSIDERANDO

Que lo actuado corresponde con las condiciones establecidas en el Reglamento de la Educación de Posgrado en la Universidad Tecnológica Nacional, Ordenanza N° 1924.

Que la solicitud cuenta con el aval de la Comisión de Posgrado de la Facultad Regional Córdoba.

Que, tratado el tema por la Comisión de Enseñanza, ésta sugiere al Consejo Directivo aprobar lo solicitado.

Por ello y atento a las atribuciones conferidas por el Estatuto Universitario en vigencia

**EL CONSEJO DIRECTIVO
DE LA FACULTAD REGIONAL CORDOBA
en su Quinta Reunión Ordinaria del día 22/10/2025
RESUELVE**

ARTICULO 1º: OTORGAR el aval académico a la propuesta de aprobación del curso de Posgrado "Machine Learning, AutoML e Interpretable Machine Learning aplicados a Minería de Datos", que obra como Anexo I de la presente resolución y que consta de diez (10) fojas.

ARTICULO 2º: ELEVAR la propuesta a la Comisión de Posgrado de la Universidad para su análisis y posterior remisión al Consejo Superior para su aprobación.

ARTICULO 3º: Regístrese, Comuníquese, Cumplido, Archívese. -

RESOLUCIÓN N°: 2034/25

Intervino
G.A.D

Ing. ROBERTO M. MUÑOZ
Secretario Académico

Ing. HÉCTOR R. MACAÑO
Decano



*Ministerio de Capital Humana
Universidad Tecnológica Nacional
Facultad Regional Córdoba*

ANEXO I RES. N° 2034/25

CURSO DE POSGRADO

**Machine Learning, AutoML e Interpretable Machine Learning
aplicados a Minería de Datos**

Duración: 60 horas

Fundamentos del recorte del campo temático

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.

Justificación.

El Machine Learning se ha consolidado como tecnología transformadora impactando todos los sectores. El AI Index Report de Stanford documenta inversión global sin precedentes en 2015-2020, generando demanda crítica de profesionales especializados.



*Ministerio de Capital Humano
Universidad Tecnológica Nacional
Facultad Regional Córdoba*

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



*Ministerio de Capital Humana
Universidad Tecnológica Nacional
Facultad Regional Córdoba*

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.

Objetivos

Que al finalizar el curso el 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

Unidades temáticas y contenidos mínimos.

Unidad 1: Fundamentos de Machine Learning y Minería de Datos

1.1. Introducción a Machine Learning: Aprendizaje supervisado, no supervisado, por refuerzo. Tipos de problemas: clasificación, regresión, clustering, reducción de dimensionalidad. Overfitting, underfitting, trade-off sesgo-varianza.



*Ministerio de Capital Humana
Universidad Tecnológica Nacional
Facultad Regional Córdoba*

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^2 . 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 previas. Neural Architecture Search (NAS) para deep learning. Estrategias de exploración del espacio de algoritmos.

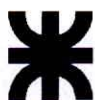


*Ministerio de Capital Humano
Universidad Tecnológica Nacional
Facultad Regional Córdoba*

- 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.



*Ministerio de Capital Humana
Universidad Tecnológica Nacional
Facultad Regional Córdoba*

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



*Ministerio de Capital Humana
Universidad Tecnológica Nacional
Facultad Regional Córdoba*

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

Bibliografía.

Fundamentos de Machine Learning

Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.

James, G., et al. (2021). An Introduction to Statistical Learning (2nd ed.). Springer.

Murphy, K. P. (2022). Probabilistic Machine Learning: An Introduction. MIT Press.

Automated Machine Learning

Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2019). Automated Machine Learning: Methods, Systems, Challenges. Springer.

Feurer, M., et al. (2015). Efficient and robust automated machine learning. Advances in Neural Information Processing Systems, 28.

Documentación oficial: Auto-Sklearn, TPOT, H2O AutoML, PyCaret, AutoGluon.

Interpretable Machine Learning

Molnar, C. (2022). Interpretable Machine Learning (2nd ed.).
[christophm.github.io/interpretable-ml-book/]

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. NIPS, 30.

Ribeiro, M. T., et al. (2016). "Why should I trust you?": Explaining predictions. ACM SIGKDD, 1135-1144.



Ministerio de Capital Humana
Universidad Tecnológica Nacional
Facultad Regional Córdoba

Minería de Datos Educativa

Romero, C., & Ventura, S. (2020). Educational Data Mining and Learning Analytics. Wiley, 10(3).

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In Learning Analytics (pp. 61-75). Springer.

Novillo Rangone, G. (2024). Automatización de un Modelo de Minería de Datos Educativa aplicando AutoML e IML [Tesis doctoral, ACOFI 2025].

Metodología de trabajo.

Metodología mixta teórico-práctica con aprendizaje experiencial:

Distribución:

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.

20% resolución de caso real propuesto por el alumno.

Estrategias pedagógicas:

Aprendizaje basado en problemas con desafíos prácticos

Trabajo colaborativo en grupos pequeños

Proyecto integrador iterativo aplicando progresivamente cada unidad



Ministerio de Capital Humana
Universidad Tecnológica Nacional
Facultad Regional Córdoba

Criterios y procedimientos de evaluación

Proyecto integrador final:

Resolver un problema real de minería de datos, preferentemente del campo profesional o interés del estudiante.

Criterios:

Rigor técnico: Aplicación correcta de metodologías, validación apropiada

Profundidad analítica: Insight generado más allá de métricas básicas

Interpretabilidad: Capacidad de explicar resultados comprensiblemente

Conciencia ética: Reflexión crítica sobre sesgos, fairness, implicaciones sociales

Comunicación: Claridad en presentación escrita y oral

Reproducibilidad: Calidad de documentación y código

Requisitos de aprobación:

Calificación mínima 7/10 en proyecto integrador

Presentación oral del proyecto final

Infraestructura y equipamiento

Se requiere que los estudiantes cuenten con una PC personal o acceso a una PC institucional e internet con software como Python 3.x, Jupyter Notebooks, Google Colab, Anaconda, y Datasets reales y públicos.

Modalidad de dictado

La modalidad de dictado será virtual síncrona. Al inicio de la cursada, se solicitará a cada estudiante un problema que le gustaría resolver para su trabajo final integrador. En caso contrario, la cátedra propondrá un sistema apropiado para su desarrollo. Las clases teóricas tendrán un enfoque expositivo, con ejemplos prácticos para resaltar los aspectos



"2025 – Año de la Reconstrucción de la Nación Argentina"

*Ministerio de Capital Humana
Universidad Tecnológica Nacional
Facultad Regional Córdoba*

teóricos. Se proporcionarán guías de ejercicios para cada unidad temática, los cuales serán resueltos por el profesor. Los alumnos podrán realizar consultas al cuerpo docente durante los encuentros virtuales. Al finalizar el curso, se ofrecerán clases de consulta específicas para resolver el trabajo integrador.

Equipo Docente: Dr. German A. Montejano

Dra. Ana G. Garis.

Dr. Gabriel A. Novillo Rangone

"2025 – Año de la Educación y el Conocimiento para una Sociedad Justa y Democratizadora"

Ing. ROBERTO M. MUÑOZ
Secretario Académico