



IV PROGRAMA DE ESPECIALIZACIÓN EN CIENCIA DE DATOS

SÍLABO

I. DATOS GENERALES

CURSO : Aprendizaje Supervisado
DURACIÓN : 16 Horas Académicas
DOCENTE : Cand. Ph.D Jordan Rodriguez Mallqui

II. SUMILLA

El curso de Aprendizaje Supervisado tiene una naturaleza teórico-práctica y brindará al participante las bases conceptuales sobre los diferentes algoritmos supervisados, así como la metodología de construcción de modelos con un enfoque de ingeniería de Machine Learning orientado a producción. Se enfatiza el uso de pipelines reproducibles, la prevención de errores comunes (data leakage), la interpretabilidad de modelos (SHAP) y la conexión entre métricas técnicas y valor de negocio.

III. COMPETENCIA FINAL

Al término del curso, el participante comprenderá en qué casos construir modelos supervisados, las consideraciones para definir la variable objetivo, el diseño de pipelines robustos, la selección e interpretación de diferentes tipos de algoritmos, y el despliegue de modelos en aplicaciones funcionales.

IV. PROGRAMA Y CONTENIDOS

SESIÓN	CONTENIDOS
1 29/11/2025	I. Ingeniería de ML y Baselines Robustos - Introducción al aprendizaje supervisado y taxonomía de algoritmos. - El "Anti-Pattern": Por qué el código de notebook no es producción. - Regresión Lineal y Regularización (Ridge, Lasso, ElasticNet). - Regresión Logística: Interpretación de coeficientes y Odds Ratios. - Pipelines de Scikit-Learn: ColumnTransformer y makepipeline. - Feature Engineering: Imputación avanzada (KNNImputer, MICE) y Target Encoding. - Data Leakage: Detección y prevención del enemigo silencioso. - Serialización de modelos con joblib. Control N°1
2 30/11/2025	II. La Revolución No-Lineal: Árboles y Ensembles - Árboles de Decisión (CART): Anatomía, Gini, y control de overfitting. - Random Forest (Bagging): La sabiduría de las masas descorrelacionadas. - Gradient Boosting: XGBoost, LightGBM y CatBoost (SOTA). - Comparativa práctica: Tiempos de entrenamiento y rendimiento. - KNN: Casos de uso modernos (imputación y recomendación). - SVM: El Kernel Trick y cuándo (no) usarlo. - Benchmarking de algoritmos en dataset real (Churn). Control N°2
3 06/12/2025	III. Optimización y Rentabilidad del Modelo - Optimización de Hiperparámetros: GridSearch vs. RandomSearch vs. Bayesiana. - Dominando Optuna: Trials, Studies y Pruning automático.



	<ul style="list-style-type: none">- Manejo de desbalance: scaleposweight vs. SMOTE.- Calibración de Probabilidades: Brier Score y CalibratedClassifierCV.- Business Analytics: Matriz de Costos y Profit Curves.- Threshold Tuning: El umbral que maximiza el dinero, no el Accuracy. Control N°3
4 07/12/2025	<p>IV. De la Caja Negra a Producción</p> <ul style="list-style-type: none">- Validación Avanzada: StratifiedKFold y TimeSeriesSplit.- Métricas finales: ROC-AUC, Precision-Recall y Lift Charts.- Explainable AI (SHAP): Interpretabilidad global y local.- Del Notebook al Script: Serialización y predict.py.- Despliegue con Streamlit: Aplicación web funcional.- Exposición de proyectos finales. Control N°4

V. METODOLOGÍA

La metodología consiste en el aprendizaje basado en proyectos con datasets reales de la industria (Credit Scoring, Customer Churn). A inicio del curso se explicará las pautas del proyecto final del curso, el cual se hará seguimiento en cada sesión. Se generarán espacios para alternar las explicaciones del Profesor con una activa participación de los alumnos mediante ejercicios prácticos en notebooks de Jupyter y talleres "hands-on" de programación en Python.

VI. SISTEMA DE EVALUACIÓN

La escala de evaluación es vigesimal. Durante el desarrollo del módulo se considerarán los siguientes criterios de evaluación:

CRITERIOS DE EVALUACIÓN	PESO
Asistencia	20%
Participación (*)	20%
Controles	40%
Examen Final	20%
TOTAL	100%

(*) Criterio de evaluación del profesor

VII. MATERIAL BIBLIOGRÁFICO

7.1. Libros, revistas

Nro.	AUTOR	TITULO	EDITORIAL
1	Géron, Aurélien	Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.)	O'Reilly Media, 2022
2	Provost, Foster & Fawcett, Tom	Data Science for Business	O'Reilly Media, 2013
3	Molnar, Christoph	Interpretable Machine Learning (2nd ed.)	Leanpub, 2022
4	Thakur, Abhishek	Approaching (Almost) Any Machine Learning Problem	Self-published, 2020
5	Zheng, Alice & Casari, Amanda	Feature Engineering for Machine Learning	O'Reilly Media, 2018

Lima, Noviembre de 2025