



Nuestro compromiso es con el *futuro*.

PROGRAMA
ESPECIALIZADO

DATA SCIENCE



DATA SCIENCE

PROGRAMA **ESPECIALIZADO**

EN ESTE PROGRAMA ESPECIALIZADO APRENDERÁS A CURAR, VISUALIZAR Y EXTRAER INFORMACIÓN DE INMENSOS VOLÚMENES DE DATOS, ADEMÁS DE ENTRENAR MODELOS PREDICTIVOS (MACHINE LEARNING) PARA LA TOMA DE DECISIÓN BASADA EN CONOCIMIENTO.

SERÁS CAPAZ DE IMPLEMENTAR DE FORMA PROFESIONAL LAS TECNOLOGÍAS MÁS NOVEDOSAS DE LA INDUSTRIA DE LA CIENCIA DE DATOS. APRENDERÁS A UTILIZAR PYTHON COMO HERRAMIENTA CENTRAL, EL LENGUAJE DE PROGRAMACIÓN ESTÁNDAR Y DE MAYOR CRECIMIENTO.

EL CURSO SE BASA EN EL CONCEPTO “APRENDER HACIENDO”: EN CUATRO SPRINTS Y UN TRABAJO INTEGRADOR CONSOLIDARÁS TODOS LOS CONOCIMIENTOS ADQUIRIDOS, IMPLEMENTANDO WORKFLOWS REALES QUE EMPLEAN LOS CIENTÍFICOS DE DATOS EN LAS EMPRESAS MÁS GRANDES DE LATAM.



CERTIFICÁ TUS CONOCIMIENTOS.
**CALIDAD GARANTIZADA CON RESPALDO
UNIVERSITARIO.**

CIENCIA DE DATOS
LA CARRERA DEL FUTURO

DIRIGIDO A

*PROFESIONALES, ESTUDIANTES Y PÚBLICO EN GENERAL
QUE DESEEN CAPACITARSE COMO UNO DE LOS
PERFILES MÁS DEMANDADOS DE LA INDUSTRIA
Y ACCEDER A FORMACIÓN DE CALIDAD.*

UN PROGRAMA ÚNICO

**HANDS-ON EXPERIENCE DURANTE TODAS LAS UNIDADES CON SPRINTS
Y UN TRABAJO INTEGRADOR FINAL**

**PODRÁS VOLVER A VER LAS CLASES DURANTE EL CURSADO,
EN EL MOMENTO QUE QUIERAS**

EJERCITARÁS EL WORKFLOW REAL DE UN CIENTÍFICO DE DATOS

ESTUDIARAS CON PROFESIONALES DE LA INDUSTRIA

**CREARÁS TU REPOSITO PERSONAL CON PROYECTOS REALES
QUE PODRÁS PRESENTAR A FUTUROS EMPLEADORES**

PROGRAMA ACADÉMICO

MÓDULO 1: INTRODUCCIÓN DATA SCIENCE

TE INTRODUCIRÁS EN EL MUNDO DE LA CIENCIA DE DATOS, APRENDIENDO CONCEPTOS INICIALES Y EL WORKFLOW DE TRABAJO REAL DE UN DATA SCIENTIST. FINALIZARÁS EL MÓDULO CAPACITADO EN EL USO DE PYTHON Y DE LAS LIBRERÍAS NUMPY, PANDAS Y MATPLOTLIB.

- **INTRODUCCIÓN:** QUÉ ES UN DATASET, TIPOS DE DATOS, TIPOS DE PROBLEMAS.
- **INTRODUCCIÓN A LA PROGRAMACIÓN CON PYTHON**
- **PROGRAMACIÓN CIENTÍFICA EN PYTHON**
 - NUMPY: LIBRERÍA DE COMPUTACIÓN CIENTÍFICA.
 - PANDAS: LIBRERÍA DE EXPLORACIÓN, CURACIÓN Y MANIPULACIÓN DE DATOS.
 - MATPLOTLIB: LIBRERÍA DE VISUALIZACIÓN DE DATOS.
- **JUPYTER NOTEBOOKS.**
- **WORKFLOW DE UN DATA SCIENTIST**

MÓDULO 2 : FEATURE ENGINEERING

IMPLEMENTARÁS TÉCNICAS AVANZADAS DE LIMPIEZA Y MANIPULACIÓN DE DATOS Y TE FAMILIARIZARÁS CON BUENAS PRÁCTICAS DE ANÁLISIS EXPLORATORIO DE DATOS (EDA).

- **FEATURE ENGINEERING CONCEPTOS:** OUTLIERS, MISSINGS, VARIABLES CATEGORICAS, DUMMIES, NUEVAS VARIABLES.
- **FEATURE ENGINEERING PANDAS:** OUTLIERS, MISSINGS, DUMMIES, NUEVAS VARIABLES.
- **FEATURE ENGINEERING SCIKIT LEARN:** OUTLIERS, MISSINGS, CATEGORICALS.

PROGRAMA ACADÉMICO

MÓDULO 3: INTRODUCCIÓN ALGORTIMOS DE APRENDIZAJE SUPERVISADO

TUS PRIMEROS PASOS EN MACHINE LEARNING. HABIENDO CONSOLIDADO CONOCIMIENTOS INICIALES EN PYTHON, COMENZARÁS A IMPLEMENTAR LA LIBRERÍA CENTRAL DE MACHINE LEARNING: SCIKIT-LEARN. COMPRENDERÁS LOS ALGORITMOS MÁS IMPORTANTES PARA APRENDIZAJE SUPERVISADO Y SUS MÉTRICAS DE VALIDACIÓN.

- **EVALUACIÓN DE MODELOS:** TRAINING/TESTING, MATRIZ DE CONFUSIÓN Y MÉTRICAS.
- **ALGORITMO KNN.**
- **ALGORITMO DECISION TREES.**
- **CLASIFICACIÓN Y REGRESIÓN CON ESTOS ALGORITMOS.**
- **OVERFITTING/UNDERFITTING.**
- **CROSS VALIDATION.**
- **PIPELINES EN SCIKIT-LEARN.**

MÓDULO 4 : SELECCIÓN DE MODELOS

APRENDERÁS A EVALUAR Y SELECCIONAR MODELOS PARA CADA PROBLEMÁTICA PARTICULAR.

- **TRADEOFF BIAS/VARIANCE.**
- **OPTIMIZACIÓN DE PARÁMETROS:** GRIDSEARCH Y HYPEROPT.
- **SELECCIÓN STEPWISE DE VARIABLES:** FORWARD, BACKWARD.
- **OTRAS MÉTRICAS:** AUC, F1, KAPPA.

PROGRAMA ACADÉMICO

MÓDULO 5: PROCESAMIENTO DE TEXTO Y ALGORITMOS DE APRENDIZAJE SUPERVISADO AVANZADOS

TRABAJARÁS SOBRE UNO DE LOS PROBLEMAS MÁS IMPORTANTES DE LA ACTUALIDAD EN LA INDUSTRIA: TEXT MINING. IMPLEMENTARÁS MODELOS DE AVANZADA COMO REDES NEURONALES, MODELOS DE ENSAMBLE BAGGING Y BOOSTING.

- **INTRODUCCIÓN A PROCESAMIENTO DE TEXTO:** BAG OF WORDS, TF-IDF, TOKENS, STEMMING.
- SVM.
- **ENSAMBLES BÁSICOS:** VOTING, PROMEDIOS.
- **ENSAMBLES BAGGING:** RANDOM FORESTS.
- **ENSAMBLES BOOSTING:** ADABOOST, XGBOOST, CATBOOST.
- **REDES NEURONALES:** PERCEPTRÓN Y BACKPROPAGATION.
- **KERAS:** REDES NEURONALES MULTICAPA.

MÓDULO 6: ALGORITMOS DE APRENDIZAJE NO SUPERVISADO Y SISTEMAS DE RECOMENDACIÓN

COMPRENDERÁS EL USO DE LOS MODELOS NO SUPERVISADOS, CENTRÁNDONOS EN SISTEMAS DE RECOMENDACIÓN: LOS ALGORITMOS IMPLEMENTADOS POR NETFLIX, SPOTIFY Y FACEBOOK, ENTRE OTRAS EMPRESAS.

- **INTRODUCCIÓN APRENDIZAJE NO SUPERVISADO.**
- **CLUSTERING: KMEDIAS, DBSCAN.**
- **REDUCCIÓN DE DIMENSIONALIDAD: PCA.**
- **INTRODUCCIÓN SISTEMAS DE RECOMENDACIÓN.**
- **REDUCCIÓN DE DIMENSIONALIDAD: SVD.**
- **SISTEMAS DE RECOMENDACIÓN FILTRO COLABORATIVO.**
- **EVALUACIÓN SISTEMAS DE RECOMENDACIÓN.**

PROGRAMA ACADÉMICO

MÓDULO 7: DEPLOY DE MODELOS

UN DIFERENCIAL DE NUESTRO PROGRAMA. APRENDERÁS A DISPONIBILIZAR MODELOS, TRABAJAR CON APIS Y DEJARLOS LISTOS PARA PRODUCCIÓN. IMPLEMENTARÁS ESTRATEGIAS QUE UTILIZAN LOS EQUIPOS DE DATOS DE LAS GRANDES EMPRESAS. CONOCERÁS LA SUITE DE IBM.

- **PERSISTENCIA DE MODELOS SCIKIT-LEARN/PYTHON.**
- **DESARROLLO API PARA ACCEDER AL MODELO Y PREDICCIONES.**
- **PUESTA EN PRODUCCIÓN CON IBM CLOUD.**

MÓDULO 8 : CÓMO SEGUIR

TE DAREMOS LINEAMIENTOS Y CONSEJOS PARA AVANZAR EN TUS CONOCIMIENTOS, ADEMÁS DE UNA INTRODUCCIÓN AL ESTADO DEL ARTE DE LA CIENCIA DE DATOS Y MACHINE LEARNING EN EL MUNDO.

- **HERRAMIENTAS CLOUD: GCP/AWS/AZURE.**
- **MACHINE LEARNING ON DEMAND: FLOYDHUB, PAPERSPACE, CRESTLE.**
- **KAGGLE Y CÓMO PERFECCIONARSE.**

4 SPRINTS DISEÑADOS A MEDIDA

CON LOS CONCEPTOS Y HERRAMIENTAS DE LOS MÓDULOS TRABAJARÁS EN SPRINTS, DONDE PONDRÁS EN PRÁCTICA LOS CONCEPTOS APRENDIDOS PARA FINALIZAR CON TU PORTAFOLIO PROFESIONAL PARA QUE USES COMO CARTA DE PRESENTACIÓN.

SPRINT I

MACHINE LEARNING DESDE CERO

COMENZARÁS A FAMILIARIZARTE CON EL ABORDAJE DE TRABAJO EN EL MUNDO DE UN DATA SCIENTIST. TRABAJARÁS CON EL LENGUAJE DE PROGRAMACIÓN PYTHON Y LAS LIBRERÍAS MÁS UTILIZADAS.

DE MANERA PRÁCTICA, ABORDARÁS LOS INTERROGANTES PRINCIPALES DE UN WORKFLOW DE CIENCIA DE DATOS:

- SI TENGO UN PROBLEMA, ¿QUÉ DATOS NECESITO PARA PODER LLEGAR A UNA POSIBLE SOLUCIÓN?
- LOS DATOS QUE REQUIERO, ¿TENGO ACCESO A ELLOS, HACE FALTA RECOLECTARLOS, CÓMO LOS PUEDO CAPTURAR?
- SI YA DISPONGO DE MI SET DE DATOS, ¿QUÉ PREGUNTAS ME PERMITE RESPONDER?

PROYECTO MI PRIMER MODELO

DADO UNA PROBLEMÁTICA TENDRÁS QUE ARMAR EL CASO DE USO, DEFINIR EL/LOS OBJETIVOS E IDENTIFICAR SI CUENTAS CON LOS DATOS REQUERIDOS PARA ESBOZAR LA SOLUCIÓN, REALIZAR UN ANÁLISIS EXPLORATORIO DEL SET DE DATOS (EDA) E IMPLEMENTAR UN MODELO SIMPLE DE MACHINE LEARNING COMO SOLUCIÓN PROPUESTA.

SPRINT II

ADVANCED MACHINE LEARNING

PROFUNDIZARÁS SOBRE LA ESTADÍSTICA Y MATEMÁTICA QUE ESTÁ DETRÁS DE CADA MODELO, PARA DAR EL SALTO HACIA MODELOS MÁS COMPLEJOS DE MACHINE LEARNING. EL OBJETIVO ES PODER HACER FOCO EN LA INTERPRETACIÓN DE RESULTADOS Y LA CORRECTA EVALUACIÓN Y COMPARACIÓN DE LOS MISMOS. AL FINALIZAR SERÁS CAPAZ DE RESPONDERNOS ACERCA DEL COMPORTAMIENTO DE NUESTRO ALGORITMO Y POR QUÉ PREDICE LO QUE PREDICE.

PROYECTO **POTENCIANDO MI SOLUCIÓN CON MACHINE LEARNING AVANZADO**

EL DESAFÍO CONSISTE EN APLICAR LAS TÉCNICAS DE INGENIERÍA DE FEATURES ENRIQUECIENDO LA INFORMACIÓN DISPONIBLE, ASÍ COMO UTILIZAR Y DESARROLLAR ALGORITMOS AVANZADOS DE MACHINE LEARNING. APLICARÁS TÉCNICAS PARA INTERPRETAR LOS RESULTADOS. RESPONDERÁS INDICANDO TUS PROPIAS CONCLUSIONES Y OPCIONES PARA MEJORAR LOS RESULTADOS.

SPRINT III

APLICACIONES MODERNAS DE MACHINE LEARNING

DURANTE ESTE SPRINT RESOLVERÁS PROBLEMAS DE DOMINIOS ESPECÍFICOS PERO DE MUCHO IMPACTO TANTO EN LA PARTE EMPRESARIAL COMO EN OTROS ÁMBITOS COMO EL CIENTÍFICO, TECNOLÓGICO, PÚBLICO, ENTRE OTROS. LOS CASOS DE USO MÁS NOVEDOSOS VAN DESDE NLP, RECOMMENDER SYSTEM, DETECCIÓN DE IMÁGENES Y TIMES SERIES CON EL OBJETIVO DE QUE PUEDAS ABORDAR CUALQUIER PROBLEMÁTICA CON AUTONOMÍA Y CAPACIDAD PARA DESARROLLAR UNA POSIBLE SOLUCIÓN.

PROYECTO MACHINE LEARNING SOLUTIONS, DESARROLLANDO APLICACIONES MODERNAS

RESOLVERÁS PROBLEMAS ACTUALES, DONDE EL MACHINE LEARNING HA MARCADO UN ANTES Y UN DESPUÉS CON SUS NOVEDOSAS SOLUCIONES. SERÁS CAPAZ DE DESARROLLAR Y UTILIZAR ALGORITMOS DE NLP, RECOMMENDER SYSTEM, TIMES SERIES Y DETECCIÓN DE IMÁGENES.

SPRINT IV

COMUNICACIÓN EFECTIVA Y PIPELINES DE DATOS

INTEGRARÁS TODOS LOS CONOCIMIENTOS APRENDIDOS, Y DARÁS EL SALTO DE UNA SOLUCIÓN INDIVIDUAL EN TU ENTORNO PERSONAL, HACIA UN AMBIENTE PRODUCTIVO EXPONIENDO TU SOLUCIÓN COMO UN SERVICIO.

PROYECTO FINAL SOLUTION, COMUNICANDO MIS RESULTADOS

NO BASTA CON TENER LA MEJOR SOLUCIÓN, NI HABER REALIZADO UN TRABAJO BRILLANTE SINO ES DISPONIBILIZADO Y COMUNICADO DE MANERA EFECTIVA CON SUS RESPECTIVAS JUSTIFICACIONES. DEBERÁS ELABORAR UN INFORME FINAL Y REALIZAR UNA PRESENTACIÓN DONDE MOSTRARÁS EL END-TO-END DE TU DESARROLLO, LOS RESULTADOS, LAS CONCLUSIONES, LAS LIMITACIONES Y LOS POSIBLES PUNTOS DE MEJORA. TAMBIÉN DEBERÁS JUSTIFICAR CÓMO TRABAJASTE CON LOS DATOS, QUÉ TRANSFORMACIONES REALIZASTE, QUÉ VARIABLES SELECCIONASTE Y POR QUÉ ELEGISTE ESE MODELO COMO TU MEJOR SOLUCIÓN, DE IGUAL MANERA QUE SE REALIZAN INFORMES SOBRE SOLUCIONES DE DATOS EN EL MUNDO EMPRESARIAL.

BONIFICACIÓN

SI SOS ESTUDIANTE DE LA UNIVERSIDAD NACIONAL DE CÓRDOBA, DE LA UNIVERSIDAD TECNOLÓGICA NACIONAL FRC O EGRESADO DE ALGÚN CURSO DE ICARO, PODÉS ACCEDER A UNA BONIFICACIÓN EXCLUSIVA DEL 5% * ADICIONAL AL ABONAR LA TOTALIDAD DEL CURSO.

**NO ACUMULABLE CON OTROS DESCUENTOS.*



ICARO Asociación Civil
CUIT 30716564815
info@icaro.org.ar
www.icaro.org.ar