

Aspectos generales

Título:	Ciencia de Datos Aplicada en Biomedicina y en la industria farmacéutica/de la salud: De la Academia a la Industria
Programas de posgrado o planes de estudio en donde se ofertará adicionalmente:	• Posgrado en Ciencias Biológicas • Posgrado en Ciencias Bioquímicas • Doctorado en Ciencias Biomédicas
Área del conocimiento:	Genética, genómica y bioinformática
Semestre:	2027-1
Modalidad:	Tópico selecto
Horario:	Martes de 9:00 am 10:30 am y Viernes de 9:00 am a 10:30 am
No. sesiones:	32
Horas por sesión:	1.5
Total alumnos PDCB:	15
Total alumnos:	25
Videoconferencia:	Si
Lugar donde se imparte:	Por videoconferencia
Informes:	insomnium098@gmail.com

Métodos de evaluación

MÉTODO	PORCENTAJE	NOTAS
Presentación	30%	
Proyecto final	70%	

Contribución de este curso/tópico en la formación del alumnado del PDCB:

Este curso contribuye a la formación del alumnado del PDCB al cerrar la brecha entre el conocimiento científico biomédico y las competencias técnicas que demanda el mercado laboral actual. Los estudiantes del programa poseen una base sólida en metodología científica, pensamiento crítico y conocimiento del dominio clínico, pero rara vez tienen la oportunidad de desarrollar habilidades prácticas de programación profesional, manejo de datos a escala o uso de herramientas estándar de la industria dentro de su formación doctoral.

Al incorporar Python (usando las buenas prácticas de la industria), control de versiones, Docker, bases de datos clínicas estandarizadas (OMOP CDM), Machine Learning, infraestructura en la nube e inteligencia artificial generativa, este curso transforma el perfil del egresado: de un científico con capacidad exclusivamente académica a un profesional versátil capaz de generar impacto tanto en la investigación como en la industria de salud, farmacéutica y tecnología. Además, las habilidades adquiridas potencian directamente su trabajo doctoral — el manejo eficiente de datos, la reproducibilidad computacional y el uso de herramientas colaborativas como GitHub elevan la calidad y el alcance de sus propios proyectos de investigación, independientemente de si su trayectoria posterior es académica o industrial.

Profesor (a) responsable

Nombre:	Martínez Gutiérrez Antonio Daniel
Teléfono:	
Email:	insomnium098@gmail.com

Profesores (as) participantes

PARTICIPANTE	ENTIDAD O ADSCRIPCIÓN	SESIONES
--------------	-----------------------	----------

MARTÍNEZ GUTIÉRREZ ANTONIO DANIEL
Responsable

Amazon EC2: Cómputo en la Nube para Ciencia de Datos
Amazon S3: Almacenamiento de Datos Clínicos con boto3
Ambientes Virtuales, Dependencias y Docker
APIs REST: Consumo de Servicios Web con Python
Aplicaciones de LLMs en Biomedicina: NER, NL2SQL y Reportes
AWS Fundamentos: IAM, Regiones y AWS CLI
Buenas Prácticas de Programación y Testing con pytest
Claude Code y Desarrollo Asistido por IA
Cohortes Avanzadas: Feasibility, Attrition Tables y Herramientas OHDSI
Construcción de APIs con FastAPI
Control de Versiones con Git y GitHub
Deployment de Modelos: FastAPI, Docker y Consideraciones Éticas
Diseño de Cohortes Clínicas en OMOP con SQL
Docker Compose y Servicios Multi-contenedor
Evaluación Clínica de Modelos: AUROC, Calibración y SHAP
Feature Engineering con Datos Clínicos
Introducción a Large Language Models y Prompt Engineering
Introducción al OMOP CDM: Arquitectura y Tablas Clínicas
Introducción: Ciencia de Datos en la Industria de Salud
Machine Learning: Fundamentos, Flujo de Trabajo y scikit-learn
Modelos Avanzados: XGBoost, LightGBM e Hiperparámetros
Navegación Jerárquica: concept_relationship y concept_ancestor
OMOP CDM: Drug Exposure, Measurement y Tablas Derivadas
Presentaciones finales
Python para Biomedicina: Estructuras de Datos y Pandas
Python para Biomedicina: Visualización y Análisis Exploratorio
RAG: Embeddings, Bases Vectoriales y Sistemas de Consulta
SQL Avanzado: CTEs, Window Functions y Subqueries
SQL Fundamentals: Queries, Joins y Agregaciones
Taller de como buscar trabajo en la industria (creación de CV, negociación de salario y aprobar las entrevistas técnicas)
Vocabularios Estándar: RxNorm y LOINC
Vocabularios Estándar: SNOMED CT e ICD-10

Introducción

La formación de los estudiantes de los posgrados de Doctorado en Ciencias Biomédicas/ Posgrado en Ciencias Biológicas y Posgrado en Ciencias Bioquímicas es reconocida por su rigor científico y por producir investigadores de alto nivel. Sin embargo, el panorama profesional de los egresados se ha diversificado enormemente en los últimos años: además de la carrera académica tradicional, hoy existe una creciente demanda en la industria farmacéutica, tecnología en salud, Real-World Evidence y ciencia de datos clínicos por profesionales que combinen pensamiento científico con habilidades técnicas de programación, manejo de datos a escala y cómputo en la nube. Este curso busca

complementar la formación científica del posgrado con las competencias prácticas que estos sectores requieren —programación profesional en Python, control de versiones, Docker, bases de datos clínicas estandarizadas (OMOP CDM), Machine Learning aplicado a salud e inteligencia artificial— para que los estudiantes cuenten con un abanico más amplio de caminos profesionales al egresar, sin dejar de ser científicos rigurosos.

La pertinencia de este curso se enmarca además en una transformación global en el manejo de datos de salud: hospitales, agencias regulatorias y compañías farmacéuticas están adoptando estándares internacionales como OMOP CDM y vocabularios como SNOMED CT y RxNorm para integrar y analizar información clínica a gran escala. Los profesionales que dominen tanto el contexto biomédico como las herramientas técnicas para trabajar con estos datos ocuparán un lugar estratégico en este ecosistema. Al completar el curso, los estudiantes habrán construido un portafolio profesional en GitHub con proyectos funcionales y documentados, tendrán experiencia práctica con las mismas herramientas que se utilizan en la industria —desde pipelines de datos y APIs hasta modelos de Machine Learning desplegados en la nube— y contarán con la capacidad de dialogar tanto con equipos clínicos como técnicos, una combinación que los hará competitivos no solo en el mercado laboral mexicano sino en posiciones remotas con empresas internacionales.

Temario

- Introducción: Ciencia de Datos en la Industria de Salud - 11 de Agosto
- Python para Biomedicina: Estructuras de Datos y Pandas - 14 de Agosto
- Python para Biomedicina: Visualización y Análisis Exploratorio - 18 de Agosto
- Control de Versiones con Git y GitHub - 21 de Agosto
- Buenas Prácticas de Programación y Testing con pytest - 21 de Agosto
- Ambientes Virtuales, Dependencias y Docker - 25 de Agosto
- Docker Compose y Servicios Multi-contenedor - 28 de Agosto
- APIs REST: Consumo de Servicios Web con Python - 1 de Septiembre
- Construcción de APIs con FastAPI - 4 de Septiembre
- SQL Fundamentals: Queries, Joins y Agregaciones - 8 de Septiembre
- SQL Avanzado: CTEs, Window Functions y Subqueries - 11 de Septiembre
- Introducción al OMOP CDM: Arquitectura y Tablas Clínicas - 18 de Septiembre
- OMOP CDM: Drug Exposure, Measurement y Tablas Derivadas - 22 de Septiembre
- Vocabularios Estándar: SNOMED CT e ICD-10 - 25 de Septiembre
- Vocabularios Estándar: RxNorm y LOINC - 29 de Septiembre

- Navegación Jerárquica: concept_relationship y concept_ancestor - 2 de Octubre
- Diseño de Cohortes Clínicas en OMOP con SQL - 6 de Octubre
- Cohortes Avanzadas: Feasibility, Attrition Tables y Herramientas OHDSI - 9 de Octubre
- AWS Fundamentos: IAM, Regiones y AWS CLI - 13 de Octubre
- Amazon S3: Almacenamiento de Datos Clínicos con boto3 - 16 de Octubre
- Amazon EC2: Cómputo en la Nube para Ciencia de Datos - 20 de Octubre
- Machine Learning: Fundamentos, Flujo de Trabajo y scikit-learn - 23 de Octubre
- Feature Engineering con Datos Clínicos - 27 de Octubre
- Modelos Avanzados: XGBoost, LightGBM e Hiperparámetros - 30 de Octubre
- Evaluación Clínica de Modelos: AUROC, Calibración y SHAP - 3 de Noviembre
- Deployment de Modelos: FastAPI, Docker y Consideraciones Éticas - 6 de Noviembre
- Introducción a Large Language Models y Prompt Engineering - 10 de Noviembre
- Aplicaciones de LLMs en Biomedicina: NER, NL2SQL y Reportes - 13 de Noviembre
- RAG: Embeddings, Bases Vectoriales y Sistemas de Consulta - 17 de Noviembre
- Claude Code y Desarrollo Asistido por IA - 20 de Noviembre
- Presentaciones finales - 24 de Noviembre
- Taller de como buscar trabajo en la industria (creacion de CV, negociación de salario y aprobar las entrevistas técnicas) - 27 de Noviembre

Bibliografía

Grus, J. (2019). Data Science from Scratch: First Principles with Python (2nd ed.). O'Reilly Media. ISBN: 978-1-492-04113-9

Wilkins, M. (2019). Learning Amazon Web Services (AWS): A Hands-On Guide to the Fundamentals of AWS Cloud. Addison-Wesley Professional. ISBN: 978-0-13-530110-4

Alammar, J., & Grootendorst, M. (2024). Hands-On Large Language Models: Language Understanding and Generation. O'Reilly Media. ISBN: 978-1-098-15095-2

Tanimura, C. (2024). SQL for data analysis. O'Reilly Media. ISBN: 978-1-098-18618-0

Kane, S. P., & Matthias, K. (2023). Docker: Up & Running: Shipping Reliable Containers in Production (3rd ed.). O'Reilly Media. ISBN: 978-1-098-13182-1

OHDSI. (2019). The Book of OHDSI: Observational Health Data Sciences and Informatics. Observational Health Data Sciences and Informatics. Disponible en: <https://ohdsi.github.io/TheBookOfOhdsi/>

Observaciones

Los alumnos deberán tener conocimientos básicos de programación en python y bash