UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE POSGRADO EN CIENCIAS BIOLÓGICAS

Denominación de la actividad académica (completa): ***Ciencia de datos en Biología***

Clave:	***Semestre:***	***Campo de conocimiento:***	***Número de Créditos:***
	2024-2	***Sistemática***	***8***

Carácter	***Horas***		***Horas por semana***	***Horas por semestre***
	Teóricas	***Prácticas***		
Optativa	***32***	***32***	***4***	***64***

Modalidad	***Duración del curso***
Curso	***Semestral***

Seriación indicativa u obligatoria antecedente, si es el caso: No aplica

Seriación indicativa u obligatoria subsecuente, si es el caso: No aplica

Objetivo general:

Que el estudiante comprenda los fundamentos básicos sobre Ciencia de datos, algoritmos de aprendizaje automático y su aplicación en datos biológicos.

Objetivos específicos:

Temario	***Horas***	
	Teóricas	***Prácticas***
Unidad 1 Introducción a la ciencia de datos *1.1. Conceptos básicos* *1.2. Programación básica* *1.3 Programación para ciencia de datos* *1.4. Colección de datos* *1.5. Calidad y limpieza de datos* *1.6. Reducción de dimensiones*	(10 hrs)	(12 hrs)
Unidad 2 Algoritmos de aprendizaje automático no supervisado *2.1. Medidas de distancia* *2.2. Agrupamiento ¡erárquico* *2.3. Agrupamiento K-means* *2.4. Reducción de dimensiones*	(6 hrs)	(6 hrs)
Unidad 3 Algoritmo de aprendizaje automático supervisado *3.1. Bosques aleatorios* *3.2. Regresión logística y clasificación* *3.3. Máquinas de vectores de soporte* *3.4. Redes neuronales*	(10 hrs)	(10 hrs)
Unidad 4 Casos de estudio *4.1 Análisis de datos de expresión genética* *4.2 Clasificación de secuencias de ADN* *4.3 Análisis de datos multi-ómicos*	(4 hrs)	(6 hrs)
Total de horas teóricas	***32***	
Total de horas prácticas	***32***	
Suma total de horas	***64***	

Bibliografía básica

Tyagi, A. K., & Abraham, A. (Eds.). (2022). Data Science for Genomics. Academic Press.
Hasija, Y., & Chakraborty, R. (2021). Hands on Data Science for Biologists Using Python. CRC Press.
Akalin, A. (2020). Computational Genomics with r. Chapman and Hall/CRC.
Yang, X. S. (2019). Introduction to algorithms for data mining and machine learning. Academic press.
Zhao, X. M. (2016). Data Mining in Systems Biology. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 13(6), 1003-1003.
Dua, S., & Chowriappa, P. (2012). Data mining for bioinformatics. CRC Press.
Chen, J. Y., & Lonardi, S. (Eds.). (2009). Biological data mining. CRC Press.
Ao, S. I. (2008). Data mining and applications in genomics (Vol. 25). Springer Science & Business Media.
Schattner, P. (2008). Genomes, browsers, and databases: data-mining tools for integrated genomic databases.

Bibliografía complementaria

Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow. " O'Reilly Media, Inc.".
Reel, P. S., Reel, S., Pearson, E., Trucco, E., & Jefferson, E. (2021). Using machine learning approaches for multi-omics data analysis: A review. Biotechnology Advances, 49, 107739.
Youens-Clark, K. (2021). Mastering python for bioinformatics. " O'Reilly Media, Inc.".
Zhou, G., Li, S., & Xia, J. (2020). Network-based approaches for multi-omics integration. Computational Methods and Data Analysis for Metabolomics, 469-487.
Grus, J. (2019). Data science from scratch: first principles with python. O'Reilly Media
Yang, X. S. (2019). Introduction to Algorithms for Data Mining and Machine Learning. Academic Press.
Lin, E., & Lane, H. Y. (2017). Machine learning and systems genomics approaches for multi-omics data. Biomarker research, 5(1), 1-6.
VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. " O'Reilly Media, Inc.".
Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). Machine learning: algorithms and applications. Crc Press.
Robert, C. (2014). Machine learning, a probabilistic perspective.
Ramsundar, B., Eastman, P., Walters, P., & Pande, V. (2019). Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more. O'Reilly Media.
Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.
Zomaya, A. Y., & Elloumi, M. (2013). Biological knowledge discovery handbook: Preprocessing, mining and postprocessing of biological data. John Wiley & Sons.
O'Neil, C., & Schutt, R. (2013). Doing data science: Straight talk from the frontline. " O'Reilly Media, Inc.".
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

Sugerencias didácticas:

__X__ *Exposición oral*
____ *Exposición audiovisual*
__X__ *Ejercicios dentro de clase*
__X__ *Ejercicios fuera del aula*
____ *Seminarios*
__X__ *Lecturas obligatorias*
__X__ *Trabajos de investigación*
____ *Prácticas de taller o laboratorio*
____ *Prácticas de campo*
____ *Otros (indicar cuáles)*

Mecanismos de evaluación del aprendizaje de los alumnos:

__X__ *Exámenes parciales*
____ *Examen final escrito*
__X__ *Tareas y trabajos fuera del aula*
____ *Exposición de seminarios por los alumnos*
__X__ *Participación en clase*
____ *Asistencia*
____ *Seminario*
__X__ *Otros (Trabajo de investigación final)*

Línea de investigación:

Perfil profesiográfico

Formación y experiencia en Bioinformática, Biología de sistemas y computación

Repartición de horas

Curso: Ciencia de datos en Biología. Sem 2024-2.

Posgrado en Ciencias Biológicas.

Clases martes y jueves de 10:00 a 12:00

Unidad 1 Introducción a la Ciencia de datos

Dr. Edgardo Galán Vásquez

11 sesiones, 22 horas

Unidad 2 Algoritmos de aprendizaje automático no supervisado

Dr. Edgardo Galán Vásquez

6 sesiones, 12 horas

Unidad 3 Algoritmo de aprendizaje automático supervisado

Dr. Edgardo Galán Vásquez

10 sesiones, 20 horas

Unidad 4 Casos de estudio

Dr. Edgardo Galán Vásquez

5 sesiones, 10 horas