



PROGRAMA OFICIAL DE CURSO (Pregrado y Posgrado)

UNIVERSIDAD DE ANTIOQUIA

1. Programa oficial del curso o microcurrículo (pregrado y posgrado)

a. Información general

Unidad Académica: Instituto de Física. Facultad de Ciencias Exactas y Naturales-

Programa académico al que pertenece: Física

Programa (s) académico (s) en los cuales se ofrece el curso: Física

Vigencia: 2024-2/2025-1

Código curso:302037

Nombre del curso: Aprendizaje estadístico

Tipo de curso: Electivo

Tipo de curso: Obligatorio

Características del curso: desplegable Validable ☒ Habilitable ☐ Clasificable ☐ Evaluación de suficiencia ☐

Modalidad educativa del curso: Presencial, asistido con TIC¹

Pre-requisitos: Probabilidad y estadística , Álgebra Lineal

Co-requisitos:

Tipo de actividades: Teórica, práctica

Número de créditos (Acuerdo 526 de marzo de 2021):

Horas de Acompañamiento directo del Docente:
64

Horas de Trabajo Independiente: 84

Horas totales del curso: 148

2. Relaciones con el perfil

Perfiles del programa académico en los que se encuentra el curso, perfiles del Egresado:

El diseño de este microcurrículo está diseñado para el siguiente perfil del estudiante:

- Esté dispuesto a trabajar en equipos interdisciplinarios en el desarrollo de proyectos de investigación básicos, aplicados o educativos, con una actitud de reconocimiento y apropiación de los aportes profesionales de otros campos del saber y de las comunidades con las cuales interactúa

Y adquiere habilidades que se adaptan al perfil de egresado:

¹ Este componente del microcurrículo se refiere a las modalidades educativas propuestas por la Vicerrectoría de la Universidad de Antioquia, las cuales son diferentes de las modalidades que presenta el Decreto 1330 y la Resolución 21795. Esta información se puede ampliar en el siguiente enlace: <https://view.genial.ly/610c3b41d209d80d7ac52752>

- Poseer la capacidad para enfrentar problemas de carácter interdisciplinario.
- Trabajar de manera autónoma y en equipos disciplinados e interdisciplinarios en la búsqueda de soluciones innovadoras y en la generación de nuevo conocimiento.
- Reconocer y utilizar fuentes de información y datos disponibles mundialmente con lo que se puedan llevar a cabo trabajos de investigación científica en su disciplina.

La ciencia de datos es un campo interdisciplinario que combina principios y prácticas de diversas áreas: estadística, matemáticas, física y computación científica, para el análisis de modelos e inferencia de los datos; ciencias de la computación, para tener un efectivo procesamiento, almacenamiento y conocimiento de las modernas arquitecturas; y expertos de dominio (ramas del saber como economía, medicina, sociología, literatura etc.) que pueden proporcionar datos y un contexto de su exploración y conocimiento subyacente. Lo que requiere una efectiva integración de los tres componentes para descubrir patrones en los datos y producir resultados significativos. Así, la ciencia de datos tiene una amplia gama de aplicaciones revolucionarias en diversas áreas, desde las ciencias naturales hasta las ciencias sociales, económicas, de la salud e ingeniería, entre otras. Por lo tanto, desarrollar habilidades de las ciencias de datos en los estudiantes de física de la Universidad de Antioquia permitirá tener profesionales más integrales con perspectivas más amplias y aplicabilidad de su saber en investigación, docencia, y/o industria. Es por ello que el propósito de este curso es buscar en el estudiante el desarrollo del trabajo interdisciplinario, matemático, estadístico y de software trabajando con problemas reales de otras áreas y problemas de las diferentes ramas de la física, buscando la independencia del estudiante, conceptualización estadística y analogía con sistemas físicos.

3. Intencionalidades formativas

Las competencias necesarias para desarrollar en ciencias de datos involucran un pensamiento computacional y algorítmico, bases de probabilidad y estadística, métodos computacionales, cálculo tensorial y álgebra lineal. El estudiante en física de la Universidad de Antioquia, en el sexto nivel, conoce todas estas competencias de los cursos de los semestres anteriores, sin embargo, no cuenta con una perspectiva algorítmica del aprendizaje estadístico y predictiva del mismo. Por eso, en este curso se busca que el estudiante desarrolle esa perspectiva a través de herramientas de ciencias de datos como: fundamentos de computación, desarrollo de software, minería de pequeños y grandes volúmenes de datos, diseño de prompts para inteligencias artificiales. Para aplicarlo en su saber y en diferentes ramas del saber, buscando el desarrollo integral, iniciativa, confiabilidad, adaptabilidad, profesionalismo del trabajo en equipo y ética del estudiante.

Objetivo General del curso:

Al terminar el curso el estudiante debe estar en la capacidad de comprender, describir, analizar, diseñar, planificar y utilizar, aplicar los conceptos de aprendizaje supervisado, no supervisado, aprendizaje por refuerzo y aprendizaje profundo en la solución de problemas básicos de pequeños y grandes volúmenes de datos en diferentes campos del saber, para promover una formación integral y de desarrollo de habilidades críticas, comunicativas, colaborativas e innovadoras en el estudiante.

Objetivos específicos.

Al final del curso el estudiante estará en la capacidad de:

- Explicar y Analizar los fundamentos teóricos de los diferentes modelos asociados a las técnicas de aprendizaje (supervisado, no supervisado, aprendizaje por refuerzo, aprendizaje profundo)
- Diseñar modelos básicos(desde cero) para las técnicas de aprendizaje e implementar modelos complejos con librerías existentes(tensor Flow, ,Scikit-learn, Keras, Pytorch etc.) para resolver problemas con datos de diferentes áreas: biología química, ciencias sociales, física, economía etc
- Diferenciar entre modelos de clasificación y predicción
- Emplear técnicas de minimización para optimizar modelos.
- Identificar el sesgo y la varianza en los datos, para analizar cómo afecta y se puede mejorar un modelo.
- Realizar exploración de los datos para identificar patrones en los datos y plantear hipótesis que pueden ser rechazadas o aceptadas.
- Evaluar las métricas de los modelos y el impacto del procesamiento de los datos.
- Crear y presentar visualización de los datos: Preprocesamiento y resultados de un modelo
- Construir pipelines de preprocesamiento de datos y entrenamiento de modelos con Python y Airflow para automatizar un modelo.
- Utilizar herramientas tecnológicas de colaboración(github, slack etc) en línea para apoyar el trabajo en equipo.
- Buscar literatura, medios audiovisuales especializadas usando herramientas contemporáneas de trabajo
- Aplicar ingeniería de prompts en modelos generativos
- Entender las operaciones MapReduce en el trabajo con grandes volúmenes
- Implementar modelos básicos en grandes volúmenes de datos

4. Aportes del curso a la formación integral e investigativa

El curso aportará la formación integral del estudiante a lo largo de los cuatro saberes: Saber, saber hacer, saber ser y saber estar. Al final del curso el estudiante estará en la capacidad de entender los conceptos fundamentales del aprendizaje estadístico y el manejo básico de grandes volúmenes de datos. Entenderá el desarrollo conceptual de los modelos asociados a las técnicas de aprendizaje, y su aplicación en la resolución de problemas de predicción y clasificación. Además, el estudiante aplicará los conceptos del preprocesamiento y visualización de datos, para elegir los mejores modelos y evaluar su rendimiento empleando herramientas de programación como Python, Tensor Flow, Keras, Airflow, pySpark entre otros. Finalmente, el estudiante será consciente de la ética y responsabilidad social al considerar cuidadosamente los sesgos y los impactos potenciales de un resultado de un modelo de aprendizaje estadístico, así también como criticar los modelos de inteligencia artificial existentes y demostrará una actitud crítica y reflexiva frente a las tecnologías contemporáneas de inteligencia artificial.

Respecto a la formación investigativa, el estudiante al final del curso estará en la capacidad de desarrollar habilidades de observación a partir de la visualización de los datos, para, plantear, validar y rechazar hipótesis, sintetizar resultados. Adicionalmente, se buscará que el estudiante realice comunicación efectiva de los resultados que se obtiene al aplicar modelos divulgando los resultados en el aula de clase. Por último, el curso busca en el estudiante promover una formación colaborativa e interdisciplinar a partir de la solución de problemas en equipo y tópicos interdisciplinarios.

5. Descripción de los conocimientos y/o saberes

Para el éxito del siguiente contenido en el aula de clase, debe de existir material de apoyo debe ser compartido con el estudiante.

Unidades	Contenido por unidad
1. introducción a la ciencia de datos	Presentación del curso
	Repaso de pandas.
	Transformación de datos con pandas
	Ejemplo de Extracción y carga de datos
	Operaciones Map, Filter, Reduce, Diseño de prompts Chat GPT, Bing
2. Modelos de Caja Negra	Modelos de clasificación y predicción derivados de los datos concepto de estimador y características
	Varianza, Bias, curvas de aprendizaje, diseño de prompts Chat GPT, Bing
3. Deducciones e implementación de Modelos	Gradiente descendente y algoritmos de minimización,
	regresión lineal, multivariada,
	técnicas de regularización, Sesgo Varianza, diseño de prompts
	Regresión Logística
	Métricas para evaluar modelos.
	Redes Neuronales, Perceptrón desde cero.
	Introducción a Keras
	Máquinas de soporte vectorial
	Árboles de decisión.
	Random Forest
	Métodos de Boosting,
	Métodos de clustering.
4. Elementos básicos de MLOPs	Procesos gaussianos, modelos probabilísticos.
	Introducción a la construcción de pipelines
5. Introducción a las redes Neuronales	Despliegue y construcción de API
	Arquitecturas básicas de redes neuronales
	Concepto de Batch, mini Batch gradiente descendente estocástico etc
	Redes Neuronales artificiales con keras y tensor flow
	Análisis de componente principales
	Análisis de discriminante Lineal
6. Introducción a los modelos generativos ética y funcionamiento.	Physics Informed
	Tokenización, Embedding, modelos generativos: Transformer.
	Ética de los modelos generativos

6. Metodología

Estrategias didácticas:

El curso se desarrollará de una forma teórico-práctica. En la primera sesión de dos horas de cada semana, el profesor explicará de forma expositiva y con ayuda de herramientas tecnológica los conceptos matemáticos y algorítmicos, buscando la participación del estudiante con el desarrollo de micro actividades auto calificables como: preguntas o desarrollos algorítmicos simples respecto a la temática que se esté tratando.

La segunda sesión de 2 horas de la semana será un laboratorio dedicado a construir nuevamente los conceptos vistos en clase y/o solución de problemas ideales, prácticos y multidisciplinarios, fomentando el desarrollo individual y el trabajo en equipo. Adicionalmente, modelos generativos y GitHub serán herramientas indispensables para mejorar la productividad en clase. Finalmente, el estudiante realizará un proyecto de manera colaborativa, en el que resolverá un problema industrial, académico o de investigación.

Medios y recursos didácticos:

- Actividades autocalificables en Jupyter.
- Laboratorios desarrollados en Jupyter.
- Ejemplos de diseños de prompts para Chat GPT, Bing
- GitHub Copilot
- Simulaciones
- Internet

Formas de interacción en los ambientes de aprendizaje y de acompañamiento del trabajo independiente del estudiante:

Respecto a los ambientes de aprendizaje será indispensable una sala de computadores para el desarrollo de las sesiones y laboratorios de clase. Los días de laboratorio, será labor del docente acompañar a los estudiantes en discusiones, para fomentar pensamiento crítico frente al desarrollo conceptual del laboratorio, haciendo preguntas que guíen al estudiante a comprender los conceptos

Tipo de problemas y Data set que se pueden trabajar en clase:

- Data set de imágenes de espectrales
- Data set de series temporales de clima
- Data set de series de tiempo experimentales en física de partículas
- Data set de estructuras atómicas
- Data set para clasificar imágenes en astronomía
- Data set de Biología, secuencias genómicas
- Data set de imágenes médicas
- Data set de experimentos físicos para aplicar *Physics informed*

Estrategias de internacionalización del currículo y del currículo en casa, que se desarrollan para cumplir con las intencionalidades formativas del micro currículo.

El objetivo central del microcurrículo se encuentra alineados con los planes internacionales de educación . Adicionalmente, durante el curso será fundamental leer en una lengua extranjera, interactuar con inteligencias artificiales en lengua extranjera y realizar lecturas de artículos en otras lenguas. Las sesiones prácticas de laboratorio incluirán dataset relacionados con problemáticas actuales como cambio climático.

7. Evaluación²

La evaluación está diseñada para fomentar en el estudiante el cuestionamiento continuo y retroalimentación rápida. El objetivo de las actividades de clase es poner en práctica lo comunicado por el profesor y los elementos básicos de la teoría, adicionalmente preguntas asociadas a las actividades también serán diseñadas para propiciar la investigación en el aula de clase. Por otro lado, el laboratorio busca la práctica de lo estudiado y comprendido en clase con material autocontenido para generar retroalimentación automática. Ejemplos de laboratorios pueden ser consultados en el siguiente link https://hernansalinas.github.io/Curso_aprendizaje_estadistico/intro.html

Momentos de evaluación	Porcentajes
Actividades (Seguimiento): Semana a Semana	10%
Laboratorios : Semana a Semana	60%
Proyecto Final : Desarrollo en el último mes	25%
Exposición Final : Finalizando el curso.	5%

8. BIBLIOGRAFÍA Y OTRAS FUENTES

Incluir solo la bibliografía que se requiere para el desarrollo del curso; además, presentar los textos en otras lenguas o traducciones que se trabajan en clase, en atención a las culturas o zonas geográficas de las que estos provienen.

Cultura o zona geográfica	Bibliografía	Palabras claves
	[1] Chambers, B., & Zaharia, M. (2018). <i>Spark: The definitive guide: Big data processing made simple</i> . O'Reilly Media.	<i>Computational and Inferential Thinking</i>

² De acuerdo con el Artículo 79 del Reglamento Estudiantil de Pregrado, “La evaluación debe ser un proceso continuo que busque no sólo apreciar las aptitudes, actitudes, conocimientos y destrezas del estudiante frente a un determinado programa académico, sino también lograr un seguimiento permanente que permita establecer el cumplimiento de los objetivos educacionales propuestos”; además en el Artículo 94 se indica que en todos los cursos se deben realizar dos o tres evaluaciones, para cumplir con las intencionalidades formativas del microcurrículo; finalmente, los artículos 95 y 96 señalan que para el desarrollo de evaluaciones parciales o finales, se pueden incluir trabajos de investigación como formas de valoración de los aprendizajes. Por su parte, en el Artículo 24 del Capítulo V del Reglamento General de Posgrados se plantea que las evaluaciones de rendimiento académico se aplicarán en todas las actividades académicas de los programas de posgrado mediante un proceso integral y transparente que permita el seguimiento al desempeño del estudiante.

	[2] Bishop, C. M. (2006). Pattern recognition and machine learning. Springer 1.	<i>Computational and Inferential Thinking</i>
	Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.	Statistical Learning
	Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference and Prediction (2nd ed.). Springer	Statistical Learning
	Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of Causal Inference: Foundations and Learning Algorithms. MIT Press.	
	R. Walpole, R. Myers, S. Myers, and K. Ye. Probabilidad y Estadística para Ingeniería y Ciencias. Pearson, New Jersey, 9na. Ed. Edition, 2012.	Probabilidad y Estadística

9. COMUNIDAD ACADÉMICA QUE PARTICIPÓ EN LA ELABORACIÓN DEL MICROCURRÍCULO

Nombres y apellidos	Unidad académica	Formación académica	Porcentaje de participación
Hernan David Salinas Jimenez	Física	Dr. En física	100%

10. APROBACIÓN DEL CONSEJO DE UNIDAD ACADÉMICA

Aprobado en Acta número del [Haga clic aquí o pulse para escribir una fecha.](#)

Nombre completo del secretario
del Consejo de la Unidad
Académica

Firma

Cargo

8. Bibliografía

- [1] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.
- [2] Bishop, C. M. (2006). Pattern recognition and machine learning. Springer 1.
- [3] Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.
- [4] Burkov, A. (2019). The Hundred-Page Machine Learning Book. Andriy Burkov 1.
- [5] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference and Prediction (2nd ed.). Springer .
- [6] https://hernansalinas.github.io/Curso_aprendizaje_estadistico/intro.html