

Programa del Curso

Información general

Profesor	Correo electrónico	Atención a estudiantes
Fabian Camilo Peña	fc.pena@uniandes.edu.co	Las citas deben solicitarse por correo electrónico.
Diego Ibagón	df.ibagon20@uniandes.edu.co	

Introducción y descripción general del curso

En los últimos años se ha experimentado un aumento considerable en el volumen y la velocidad con la que se generan nuevos datos. Este fenómeno es consecuencia del incremento en la cantidad de dispositivos electrónicos conectados así como de las capacidades de procesamiento y almacenamiento de datos a las que hoy tenemos acceso. Hoy en día, los datos son un activo más de las organizaciones y como con cualquier otro activo, es propicio identificar y explotar diferentes maneras de extracción de valor. La ciencia de datos aplicada y en particular, el rol del científico de datos en las organizaciones, es el encargado de identificar y explotar estas oportunidades.

Los datos pueden generarse desde múltiples fuentes, entre ellas, procesos de negocio que involucran operaciones entre colaboradores, clientes, proveedores, entre muchos otros actores. Algunas de las responsabilidades del científico de datos son extraer datos de esas fuentes, integrarlos, transformarlos, analizarlos y entregarlos, utilizando diferentes herramientas y técnicas computacionales y de estadística, de forma tal que puedan ser utilizados para la toma de decisiones. La utilidad de los datos se define en términos de los beneficios que reciban las organizaciones con su uso, por ejemplo, al adquirir nuevos clientes, disminuir costos, estimación de la demanda, optimización de procesos de negocio, automatización de tareas, entre muchos otros usos.

Este es el marco de trabajo de este curso, el cual pretende principalmente desarrollar habilidades en los estudiantes en la selección y uso de técnicas y herramientas computacionales y de estadística, apropiadas para el análisis de datos, incluyendo formas efectivas de transmitir los resultados o hallazgos de dichos análisis.

Objetivos pedagógicos

- Entender los fundamentos de la ciencia de datos, el rol del científico de datos en las organizaciones y las metodologías actuales para el desarrollo de este tipo de proyectos de analítica de datos.
- Desarrollar habilidades en el uso de herramientas y técnicas apropiadas para preparar y analizar datos a pequeña y mediana escala.

- Preparar y realizar análisis de diferentes fuentes de datos con el fin de extraer información útil y accionable para una organización.
- Utilizar de forma apropiada técnicas computacionales basadas en estadística y aprendizaje automático para la automatización de toma de decisiones.
- Comunicar efectivamente los resultados o hallazgos de un proyecto de ciencia de datos a los *stakeholders* del negocio, habilitando la toma de decisiones con base en datos.
- Entender y aplicar mecanismos básicos de despliegue u operacionalización de modelos basados en aprendizaje automático para que puedan ser integrados efectivamente con los demás procesos de la organización.

Plan de temas

1. Generalidades: Ciencia de datos, el rol del científico de datos, metodologías (2 sesiones).
2. Análisis Exploratorio de Datos (EDA), visualización de datos, inferencia estadística (2 sesiones).
3. Limpieza y preparación de datos (1 sesión).
4. Analítica predictiva, aprendizaje supervisado:
 - a. Regresión lineal, evaluación, interpretación, regularización (2 sesiones).
 - b. Clasificación, algoritmos, evaluación, selección de modelo, explicabilidad (3 sesiones).
5. Despliegue de modelos, MLOps, gobierno de datos (2 sesiones).
6. Proyecto: Seguimiento y evaluación (2 sesiones).

Metodología

El curso se desarrolla en torno a casos de estudio los cuales deben ser desarrollados utilizando como guía el material proporcionado durante las sesiones de clase así como la bibliografía sugerida. Se proponen sesiones de clases teóricas complementadas con talleres prácticos que permitan apropiar los diferentes conceptos, habilidades y herramientas que son objeto de aprendizaje del curso. Los estudiantes deben preparar previamente cada sesión con el fin de realizar discusiones que permitan profundizar en los temas vistos así como avanzar en los talleres basados en casos propuestos durante la clase. Los talleres se trabajan usando Python como lenguaje de programación en el ambiente Jupyter Labs (Anaconda o Google Colab). Los dos primeros talleres se deben realizar de forma individual y los siguientes de forma grupal de máximo 3 integrantes. El proyecto final integra los conceptos y habilidades vistos en clase y se deben trabajar en grupos de máximo 3 integrantes.

Evaluación y aspectos académicos

Generalidades

- Clases: **3** horas semanales con participación activa de los estudiantes.
- Canales oficiales de comunicación: el correo electrónico uniandes de los profesores, lista de correo del curso, sistema de apoyo a la docencia BloqueNeón (<http://bloqueneon.uniandes.edu.co>) y página web del plan de estudios de MINE, en donde encuentra la información general del curso (<https://sistemas.uniandes.edu.co/es/mine>).

Evaluación del curso

Quices	5 quices (2% c/u)	10%
Talleres	2 talleres individuales (9% c/u), 2 talleres grupales (6% c/u)	30%
Parciales prácticos	2 parciales (15% c/u)	30%
Proyecto		30%
	Entrega 1: Análisis descriptivo y avance de regresión (si aplica).	5%
	Entrega 2. Análisis predictivo, regresión y/o clasificación.	10%
	Entrega 3. Análisis completo, despliegue, presentación de resultados en términos de negocio (sustentación).	15%

Política de aproximación de notas finales

- Las notas del curso varían entre 0.00 y 5.00 con dos decimales sin aproximaciones.
- Para pasar el curso es indispensable lograr un puntaje ponderado 3.00/5.00.
- No existe aproximación automática en la nota definitiva. En particular, no hay aproximación a 3.00 de puntajes menores a esta nota (e.g., 2.99 no es 3.00).

Protocolo MAAD

El miembro de la comunidad que sea sujeto, presencie o tenga conocimiento de una conducta de maltrato, acoso, amenaza, discriminación, violencia sexual o de género (MAAD) deberá poner el caso en conocimiento de la Universidad. Esto con el propósito de que se puedan tomar acciones institucionales para darle manejo adecuado al caso, a la luz de lo previsto en el protocolo y velando por el bienestar de las personas afectadas.

Para poner en conocimiento el caso y recibir apoyo, usted puede contactar a:

1. **Línea MAAD:** lineamaad@uniandes.edu.co
2. **Ombudsperson:** ombudsperson@uniandes.edu.co
3. **Decanatura de Estudiantes:** centrodeapoyo@uniandes.edu.co
4. **Red de Estudiantes:**
 - PACA (Pares de Acompañamiento contra el Acoso) paca@uniandes.edu.co
5. **Consejo Estudiantil Uniandino (CEU)** comiteacosoceu@uniandes.edu.co

Política de momentos difíciles

Estamos viviendo situaciones complejas de salud pública y entorno social, en el mundo y particularmente en nuestro país. Por ello, la Universidad brinda a toda la comunidad apoyo en muchas dimensiones: emocional, de salud y tecnológico para la virtualidad. Existen líneas directas de atención a través de los canales <https://virtualidad.uniandes.edu.co/>, virtualidad@uniandes.edu.co, <https://uniandes.edu.co/es/soy/servicios/salud> y <https://decanaturadeestudiantes.uniandes.edu.co/images/stories/Imagenes/2020-10/apoyo/960/lineas-de-apoyo.jpg>.

Todas las personas pueden pasar por un momento difícil que de alguna manera pueda afectar nuestra vida en la Universidad. Pueden ser problemas en casa, con la pareja, incluso estrés por esta u otra materia. Si usted siente que está pasando por un momento complicado, sin importar el motivo, siéntase con la tranquilidad de

hablar con alguno de los profesores para pedir tiempo o apoyo. Ningún trabajo o entrega puede sobrepasar su salud mental y física. Su bienestar es lo más importante.

Bibliografía disponible en la biblioteca de la Universidad

Fundamentos

- [1] Cathy O'Neil and Rachel Schutt. Doing Data Science: Straight Talk from the Frontline. O'Reilly Media, Inc. 2013.
- [2] Daniel Vaughan. Data Science: The Hard Parts. O'Reilly Media, Inc. 2024.
- [3] Foster Provost and Tom Fawcett. Data Science for Business. O'Reilly Media, Inc., 2013.
- [4] Doug Rose. Data Science: Create Teams That Ask the Right Questions and Deliver Real Value. APress. 2016.

Estadística

- [5] Peter Bruce, Andrew Bruce and Peter Gedeck. Practical Statistics for Data Scientists, 2nd Edition. O'Reilly Media, Inc. 2020.
- [6] 365 Careers Ltda. Statistics for Data Science and Business Analysis. O'Reilly Media, Inc., 2021.

Visualización

- [7] Cole Nussbaumer Knaflic. Storytelling with Data: A Data Visualization Guide for Business Professionals. Wiley Publishing. 2015.

Machine Learning

- [8] Nathan George. Practical Data Science with Python. O'Reilly Media, Inc. 2021.
- [9] Aurélien Géron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd Edition. O'Reilly Media, Inc. 2022.
- [10] Christoph Molnar. Interpretable Machine Learning: Making Black Box Models Explainable, 2nd Edition. 2022.
- [11] Valliappa Lakshmanan, Sara Robinson and Michael Munn. Machine Learning Design Patterns. O'Reilly Media, Inc. 2020.
- [12] Chip Huyen. Designing Machine Learning Systems. O'Reilly Media, Inc. 2022.