

**FACULTAD / ESCUELA DE  
DEPARTAMENTO DE*****Información de la asignatura***

<b>Nombre de la asignatura</b>	Fundamentos de Analítica 2
<b>Código de la asignatura</b>	60156
<b>Período académico</b>	2023-1
<b>Intensidad total</b>	52 horas
<b>Créditos<sup>1</sup></b>	4
<b>Docente(s)</b>	Javier Diaz Cely – <a href="mailto:jdiazcely@yahoo.com">jdiazcely@yahoo.com</a> Diego Fernando Agudelo – <a href="mailto:diegoagudelo30@gmail.com">diegoagudelo30@gmail.com</a>

***Introducción o presentación general del curso***

Este curso continúa la exploración de los métodos de analítica de datos iniciada en el curso Fundamentos de Analítica 1. Se centra en dos grandes bloques.

En una primera instancia, utilizando el lenguaje de programación Python, se estudiarán modelos estadísticos de series de tiempo que permiten pronosticar el comportamiento futuro de un fenómeno. Estos modelos consideran observaciones de la misma variable objetivo en periodos anteriores para predecir su comportamiento futuro. Entre los modelos estudiados tenemos los de suavización, los auto regresivos y los de media móvil. Se abordarán además particularidades de estos modelos, como las métricas y protocolos a seguir para poder evaluar la calidad predictiva de los modelos creados.

En una segunda instancia, se continuará la exploración de modelos de aprendizaje automático no tratados en el primer curso, como los modelos de máquinas de vectores de soporte y de redes neuronales artificiales, en tareas como la clasificación, la regresión, la detección de anomalías y el pronóstico. Cabe anotar que en esta segunda instancia se utilizará el lenguaje de programación Python, con mayores funcionalidades en cuanto a las técnicas tratadas.

***Formación en competencias***

Competencias de la Maestría en Ciencia de Datos trabajadas en este curso.

**C: Competencia / RA: Resultado de Aprendizaje**

**C3. Solución de Problemas.** Solucionar problemas complejos usando pensamiento sistémico e información relevante.

**C3RA1.** Examinar problemas complejos aplicando el pensamiento sistémico y plantear soluciones basadas en conocimiento e información relevante.

**C4. Análisis de información:** Utilizar análisis de datos, técnicas estadísticas y aprendizaje automático sobre datos e información disponible para descubrir relaciones y obtener un mejor entendimiento de los problemas y los procesos organizacionales, con el fin de respaldar la toma de decisiones.

**C4RA1.** Examinar conjuntos de datos aplicando técnicas estadísticas y aprendizaje automático con el fin de tomar decisiones informadas soportadas en datos.

### ***Objetivo general–meta de aprendizaje***

Al finalizar el curso, los estudiantes podrán de aplicar modelos de pronóstico, de máquinas de vectores de soporte y de redes neuronales artificiales a problemáticas reales que permitan crear valor en las organizaciones.

### ***Objetivos terminales–resultados de aprendizajes de la asignatura***

- OT1: Identificar problemáticas de negocio que requieran la aplicación de modelos de pronóstico, proponiendo soluciones a partir del desarrollo de modelos predictivos
- OT2: Reconocer las diferencias entre los diferentes modelos de pronóstico, tanto los basados en estadística, como los basados en aprendizaje automático
- OT3: Aplicar Python para el entrenamiento y evaluación de modelos de clasificación, regresión y pronóstico con datos no estructurados.
- OT4: Construir modelos de Deep Learning utilizando frameworks de Python.

<b>Objetivo terminal–resultado de aprendizaje del curso o asignatura</b>	<b>Competencia en formación</b>	<b>Resultado de aprendizaje de la competencia de egreso al que se contribuye</b>
OT1	C3	C3RA1
OT2	C3	C3RA1
OT2	C4	C4RA1
OT3	C4	C4RA1
OT4	C4	C4RA1

### ***Unidades de aprendizaje***

De acuerdo con el desarrollo del curso o asignatura, detalle cada unidad de aprendizaje con sus respectivos objetivos específicos de aprendizaje (saberes conceptuales, procedimentales y/o actitudinales), contenidos temáticos, proyectos a realizar, rotación, técnicas y/o enfoques metodológicos. Los objetivos específicos de aprendizaje deben estar en articulación con los objetivos terminales-resultados de aprendizaje del curso.

#### **Unidad de aprendizaje 1: Modelos de pronóstico basadas en series de tiempo**

##### **Objetivos específicos de aprendizaje**

1. Entender las características generales de los problemas de pronóstico y de las series de tiempo (tipos de datos, métricas, protocolos de evaluación)
2. Descomponer una serie de tiempo en sus diferentes componentes (tendencia, estacionalidad, error)
3. Analizar la estacionariedad de una serie de tiempo y reconocer sus implicaciones en los modelos de pronóstico
4. Aplicar modelos de pronóstico de suavizamiento, auto regresivos, de media móvil, ARIMA

## 5. Aplicar el lenguaje Python en series de tiempo

### Unidad de aprendizaje 2: Modelos de aprendizaje supervisado

#### Objetivos específicos de aprendizaje

1. Aplicar el lenguaje Python en procesos de analítica de datos
2. Distinguir entre modelos discriminantes y modelos generativos de aprendizaje automático.
3. Desarrollar modelos de máquinas de vectores de soporte para tareas de clasificación, regresión y detección de anomalías, utilizando el lenguaje Python.
4. Desarrollar modelos de redes neuronales artificiales para tareas de clasificación y regresión, utilizando el lenguaje Python y el framework PyTorch.
5. Desarrollar modelos de redes neuronales convolucionales, utilizando un framework de Deep Learning en el lenguaje Python y el framework PyTorch.

### Metodologías de aprendizajes

El curso se desarrollará por unidades, de acuerdo con el contenido presentado, con espacios de discusión, aplicación y análisis de los conceptos, y la participación activa de los estudiantes.

Los estudiantes deberán preparar, antes de la clase, los temas que asigne el profesor. Bajo el esquema de trabajo de este curso, preparar un tema significa hacer una lectura crítica (análisis y síntesis) del tema/material de lectura que corresponda, indagar sobre los aspectos desconocidos, resolver las preguntas y los ejercicios planteados, y llegar a clase dispuesto a discutir el tema y a resolver las dudas que hayan surgido al realizar las actividades mencionadas y las propuestas por el profesor.

### Evaluación de aprendizajes

Las dos unidades tienen una participación proporcional al número de sesiones. La primera unidad, que consta de 4 clases, tiene un peso de  $4/11 = 36\%$  de la nota. La segunda unidad, que consta de 7 clases, tiene un peso de  $7/11 = 64\%$  de la nota final.

Mecanismo o actividad evaluativa	Porcentaje de la nota final	Relación con objetivo terminales – resultado de aprendizaje del curso	Relación con el resultado de aprendizaje de la competencia de egreso
Quices unidad 1	6%	OT1, OT2	C3RA1, C4RA1
Taller unidad 1	10%	OT1, OT2	C3RA1, C4RA1
Examen unidad 1	20%	OT1, OT2	C3RA1, C4RA1
Quices unidad 2	17%	OT1, OT3, OT4	C1RA1, C3RA1, C4RA1
Taller 1 unidad 2	13.5%	OT1, OT3, OT4	C1RA1, C3RA1, C4RA1
Taller 2 unidad 2	13.5%	OT1, OT3, OT4	C1RA1, C3RA1, C4RA1
Examen unidad 2	20%	OT1, OT3, OT4	C1RA1, C3RA1, C4RA1

Los exámenes del curso comprenden los temas en sus aspectos tanto teóricos (en forma de cuestionario de escogencia múltiple, incluyendo además pequeños ejercicios) como prácticos

(caso de aplicación), y el profesor podrá eventualmente realizar evaluaciones cortas con o sin previo aviso, orales o escritas. En el caso de las sustentaciones de trabajos grupales, serán tanto individuales como grupales.

El trabajo de aplicación se hará sobre un dataset escogido por el profesor o sobre uno propuesto por los estudiantes, que deberá ser validado previamente por el profesor.

## ***Medios Educativos.***

### **GENERAL**

1. Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python (capítulo 11), Sebastian Raschka et al., 2022 Packt
2. Deep Learning with PyTorch, Eli Stevens et al, 2020 Manning
3. An Introduction to Statistical Learning with Applications in R. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. Springer, 2009. Gratis on-line
4. Deep Learning with Python, 2nd edition. Francois Chollet. Manning, 2020
5. Notas de clase

**Software de referencia:** Todas las aplicaciones utilizadas durante el curso son de acceso libre.

- Python Scikit-learn <https://scikit-learn.org/stable/>
- Anaconda <https://www.anaconda.com/download/>
- Jupyter Notebook <https://jupyter.org/>
- Google Colab: <https://colab.research.google.com/>