

UNIVERSIDAD NACIONAL DE ASUNCIÓN
FACULTAD POLITÉCNICA
LICENCIATURA EN CIENCIAS INFORMATICAS
ENFASIS EN ANALISIS DE SISTEMAS INFORMATICOS
PLAN 2009
PROGRAMA DE ESTUDIOS

I. - IDENTIFICACIÓN

1.	Asignatura	: Electiva III – Ciencia de Datos - Data Science
2.	Grupo	: Siete
3.	Horas semanales	: 5 horas
3.1.	Clases teóricas	: 2 horas
3.2.	Clases prácticas	: 3 horas
4.	Total real de horas disponibles	: 90 horas
4.1.	Clases teóricas	: 36 horas
4.2.	Clases prácticas	: 54 horas

II. - JUSTIFICACIÓN

A medida que nuestra economía, la sociedad y la vida diaria se vuelven cada vez más dependientes de los datos. Así, la ciencia de datos emerge como un campo multifacético utilizado para obtener información a partir de datos complejos.

La Ciencia de datos se puede definir como una nueva “trans-disciplina” que se consolida mediante la síntesis de un relevante número de disciplinas y conocimientos incluidos la computación, las estadísticas, la informática, comunicaciones, administración y la sociología para estudiar cómo la enorme cantidad de datos hoy disponibles, pueden convertirse en nuevos conocimientos para dominios bien específicos como, la medicina, los negocios, las ciencias políticas y cualquier otra área en la que se genera gran cantidad de datos que hoy no son suficientemente explotados

Adquirir habilidades proporcionadas por la ciencia de datos es esencial para los estudiantes universitarios actualmente y en el futuro para formar profesionales con competencias destacadas en investigación y aplicación en el área de Ciencia de datos para abordar nuevos desafíos en el manejo de la información en diversos dominios de conocimiento.

La evolución de la ciencia de datos significa oportunidades para la investigación de vanguardia, la innovación tecnológica y una nueva economía de datos

III. - OBJETIVOS

- Generar nuevos conocimientos mediante la obtención y recopilación de información valiosa utilizando las matemáticas y técnicas informáticas inteligentes para tomar decisiones o resolver problemas de manera eficiente.
- Desarrollar soluciones a problemas de datos en una amplia gama de entornos, como logística, medicina, agricultura, pecuaria, robótica entre otros.
- Investigar en diversos dominios del conocimiento para contribuir con investigación e innovación
- Desarrollar habilidades para diseñar, preparar, analizar y manejar grandes volúmenes de información, estructurada y no-estructurada.
- Planificar una vinculación efectiva entre industria y academia para fomentar la investigación y el desarrollo de tecnologías de análisis y explotación de grandes datos en el país.
- Manifiestar compromiso en la formación de lider y emprendedor en tecnologia

IV. - PRE-REQUISITO

- Ingeniería de Software I
- Base de Datos III

V. - CONTENIDO

5.1. Unidades programáticas

1. Introducción a la Ciencia de datos.
2. Programación con Python para Ciencia de datos.
3. Análisis exploratorio de datos.
4. Ingeniería de datos.
5. Machine Learning.
6. Visualización de datos y comunicación de los resultados.
7. Recuperación de la información desde repositorios de datos masivos.
8. Cuestiones Legales en el manejo de los datos.
9. Presentación de un Proyecto Final.

5.2. Desarrollo de las unidades programáticas

1. **Introducción a la Ciencia de datos**
 - 1.1. Visión general completa de los temas centrales en Ciencia de Datos, desde el punto de vista matemático y desde la perspectiva de la Inteligencia Artificial.
 - 1.2. Clases básicas de técnicas y métodos

- 1.3. Fundamentos teóricos de la ciencia de datos y la inteligencia computacional.
- 1.4. Principales dominios de aplicación de la ingeniería del conocimiento.
- 2. Programación con Python para Ciencia de datos**
 - 2.1. Conceptos de programación y aplicaciones usando el lenguaje Python.
 - 2.2. Introducción a las bibliotecas esenciales para Ciencia de datos.
- 3. Análisis exploratorio de datos**
 - 3.1. Técnicas exploratorias esenciales para la comprensión de los datos que se aplican antes de comenzar el modelado formal para ayudar en la percepción de los modelos estadísticos más complejos.
 - 3.2. Técnicas exploratorias importantes para eliminar o agudizar las hipótesis potenciales a ser abordadas por los datos.
- 4. Ingeniería de datos**
 - 4.1. Manejo de los datos crudos.
 - 4.2. Técnicas de depuración, limpieza y preparación.
- 5. Machine Learning**
 - 5.1. Técnicas básicas y avanzadas de aprendizaje automático.
 - 5.2. Uso práctico de técnicas y problemas para desarrollar aplicaciones reales de aprendizaje automático.
 - 5.3. Prácticas de laboratorio para comprender el funcionamiento de los algoritmos de aprendizaje automático.
- 6. Visualización de datos y comunicación de los resultados**
 - 6.1. Exploración de datos en forma visual.
 - 6.2. construcción de un relato para comunicar los resultados obtenidos del análisis de los datos.
- 7. Recuperación de la información desde repositorios de datos masivos**
 - 7.1. Fundamentos de la recuperación de información, su importancia.
 - 7.2. Casos prácticos.
- 8. Cuestiones Legales en el manejo de los datos**
 - 8.1. Normativas que se aplican dentro del ámbito de la Ciencia de Datos a nivel nacional e internacional.
 - 8.2. Límites que se imponen y obligaciones que conlleva el manejo de datos masivos.
- 9. Presentación de un Proyecto Final**

VI. - ESTRATEGIAS METODOLÓGICAS

1. Exposición de fundamentos teóricos con apoyo audiovisual.
2. Trabajos de investigación y evaluación de herramientas y técnicas a desarrollarse fuera del horario de clases.
3. Trabajo de diseño y programación de un proyecto de Ciencia de Datos
4. Evaluación de trabajos en laboratorio

VII. - MEDIOS AUXILIARES

1. Pizarras acrílicas
2. Marcadores
3. Borrador de pizarra acrílica
4. Computadores
5. Equipo multimedia
6. Plataforma Virtual EDUCA
7. Sala de laboratorio equipada para las prácticas.
8. Acceso a internet

VIII. - EVALUACIÓN

Para evaluar la asignatura se tienen en cuenta lo siguiente:

1. Examen parcial teórico de 30% sobre el total.
2. Examen final de teoría y práctica de 40% sobre el total.
3. Trabajo practico de 20% del total
4. Evaluación de ejercicios de laboratorio de 10% sobre el total.

IX. - BIBLIOGRAFÍA

- Baumer, B. (2015). *A data science course for undergraduates: Thinking with data. The American Statistician* 69.4. 334-342.
- Cao, L. (2017). *Data science: challenges and directions. Communications of the ACM* 60.8 (2017): 59-68.
- Cover, T. M. & Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.
- Hastie, T., & Tibshirani, R. & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Kelleher, J. D. & Tierney, B. (2018). *Data science*. MIT Press.
- Larose, D. T. (2015). *Data mining and predictive analytics*. John Wiley & Sons.
- VanderPlas, J. (2016). *Python data science handbook: essential tools for working with data*. O'Reilly Media, Inc.

ENLACES WEB

- An Introduction to data science, Part 1: Data, structure and the data science pipeline. Disponible en: <https://www.ibm.com/developerworks/library/ba-intro-data-science-1/ba-intro-data-science-1-pdf.pdf>
- Python™. Disponible en: <https://www.python.org/>

COLECCIÓN DE LA BIBLIOTECA DE LA FACULTAD POLITÉCNICA

- Daza Vergaray, A. (2017) Data minig = Minería de datos. Lima: Macro EIRL
- Larose, D. T. & Larose, C. D. (2018). Data mining and predictive analytics. Nueva Delhi: Wiley India

BASE DE DATOS ON LINE

- Kotu, V., & Deshpande, B. (2019). *Data science : Concepts and practice* (Second ed.) [Second edition.]. Cambridge, MA: Morgan Kaufmann, an imprint of Elsevier. (2019). Disponible en: CICCO Conacyt Paraguay