



ANEXO 1

Diplomatura Universitaria en Ciencia de Datos y Python

Introducción

En esta Diplomatura, aprenderemos cómo usar Python para ciencia de datos: recopilación, limpieza, visualizaciones y construcción de modelos descriptivos y predictivos.

El objetivo principal es desarrollar primeras habilidades analíticas y de programación para análisis de datos. Para lograrlo, vamos a resolver situaciones problemáticas aplicadas y proyectos en los cuales se pondrán en práctica todos los conceptos de programación utilizados en ciencia de datos.

La herramienta de aplicación será el lenguaje Python con las principales librerías que se usan habitualmente en este campo de estudio, como Pandas, Numpy y Scikit Learn.

Nuestro horizonte es que en la era de la información, los datos nos rodean por todas partes y en volúmenes inmensos. Dentro de estos datos hay respuestas a preguntas de diferentes sectores sociales (política, negocios, ciencia, etc.). Seguro que teniendo acceso a un gran conjunto de datos y pudiéndolos analizar, seríamos capaces de encontrar las respuestas a muchos problemas.

Fundamentación

Python es un lenguaje multiplataforma que puede ser programado mediante Programación Orientada a Objetos. Su sintaxis es simple, distinguiéndose de otros lenguajes como Java, C, PHP y JavaScript en la ausencia de paréntesis y la no incorporación de punto y coma al final de cada instrucción, entre otros aspectos. Es un lenguaje muy completo, con soporte para bases de datos relacionales, no relacionales u orientados a objetos, capaz de incorporar ORM's, con frameworks de desarrollos complejos.

En este trayecto formativo sentaremos las bases del lenguaje aprendiendo a trabajar con sus diferentes tipos de objetos y bases de datos.

Se espera que los participantes incorporen criterios para diseñar y utilizar repositorios de datos.

Objetivos generales

Formar profesionales en Ciencia de Datos que:

- puedan adaptarse a diferentes contextos de disciplinas, servicios y trabajo
- comprendan y puedan aplicar estructuras básicas de la programación, en particular orientada al análisis de bases de datos, usando Python como lenguaje de programación.



- Puedan diseñar un repositorio de datos.
- Puedan reconocer la pertinencia de herramientas del análisis estadístico

¿Qué se espera aprender?

- Procedimientos básicos en un proyecto que implique herramientas de ciencia de datos.
- Programar usando scripts de Python y Jupyter notebooks
- Cómo manipular y analizar conjuntos de datos.
- Realizar análisis estadísticos básicos e implementar métodos de aprendizaje automático.
- Limpieza, depuración y análisis de conjuntos de datos

Destinatarios

Esta diplomatura está dirigida a todos los interesados en ingresar al mundo de la ciencia de datos. Es conveniente una experiencia previa con cualquier lenguaje de programación (Java, C, Pascal, Fortran, C ++, R, Python, PHP, etc.).

Duración y modalidad

- 4 meses de cursado.
- Cantidad de horas: 120 hs. reloj Teórico-Prácticas, incluyendo un Proyecto Final Integrador.
- Modalidad de cursado virtual, mediante plataforma educativa Moodle (actividades sincrónicas y asincrónicas).

Lugar

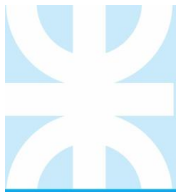
El ámbito de coordinación y supervisión, preparación de materiales didácticos, recursos tecnológicos, construcción y gestión de las aulas virtuales y de las plataformas de simulación on-line, será en la Facultad Regional Concordia, al igual que los encuentros que sean presenciales.

Cupo

El cupo máximo por cada cohorte será de 50 (cincuenta) participantes. Podrá habilitarse varias cohortes simultáneamente, dependiendo de la disponibilidad de docentes y de los recursos necesarios.

Requisitos de admisión

Escuela secundaria completa



Plan de estudios

Módulo 1: Introducción a la programación con Python

Introducción a Python y su sintaxis básica
Estructuras de control de flujo
Funciones y módulos en Python
Programación orientada a objetos en Python

Módulo 2: Fundamentos de estadística y probabilidad.

Conceptos básicos de estadística: descriptiva e inferencial. Medidas de centralización y dispersión. Representaciones gráficas.

Probabilidad: conceptos básicos, distribuciones de probabilidad. Simulación de experimentos aleatorios. Introducción a los Métodos estadísticos.

Modelización estadística: regresión lineal, regresión logística.

Módulo 3: Manipulación de datos con Python

Manejo de datos en Python: estructuras de datos, archivos. Introducción a SQL.

Librerías para manipulación de datos: NumPy, Pandas.

Exploración y visualización de datos con Python: Matplotlib, Seaborn.

Módulo 4: Aprendizaje automático y minería de datos

Introducción al aprendizaje automático y minería de datos

Técnicas de pre procesamiento de datos: limpieza, normalización, transformación.

Técnicas de aprendizaje no supervisado.

Introducción a Tableau, QlikSense, Splunk, QlikView

Módulo 5: Proyecto de ciencia de datos con Python

Desarrollo de un proyecto de ciencia de datos utilizando Python y las técnicas aprendidas en los módulos anteriores. El proyecto puede ser seleccionado por los estudiantes o propuesto por el docente.

Este diseño está enfocado en proporcionar una base sólida en programación con Python, estadística y manipulación de datos, así como en técnicas de aprendizaje automático y minería de datos. El proyecto final busca que los estudiantes apliquen los conocimientos adquiridos en una tarea práctica y real, permitiéndoles desarrollar habilidades en resolución de problemas y trabajo en equipo.



Planificación

Inicio: 16 de agosto de 2023

Equipo docente: Carlos Casanova - Mario Álvarez - Juan Carlos Teze - Mario González – Carlos Blanc

Fecha aproximada	Tema	Área
16/08	Introducción a Python y su sintaxis básica	Python
23/08	Estructuras de control de flujo. Funciones y módulos en Python	Python
30/08	Estructuras de datos básicas (lista, diccionario). Programación orientada a objetos en Python	Python
06/09	Conceptos básicos de estadística: generalidades, población, muestras, variables. Tipos de estudios	Estadística
13/09	Tipos de variables: representaciones tabulares y gráficas.	Estadística
20/09	Descripción numérica de un conjunto. Casos de estudio: datos univariados y bivariados.	Estadística
27/09	Nociones básicas de probabilidad, variable aleatoria, esperanza, varianza, convergencias.	Estadística
04/10	Simulación de experimentos aleatorios: distribuciones muestrales, nociones de estimación puntual y por intervalos.	Estadística
11/10	Modelización estadística: nociones de ajuste, métricas de bondad del ajuste, hipótesis estadísticas.	Estadística
18/10	Modelización estadística: ejemplos (regresión lineal, regresión logística, casos de estudio)	Estadística
25/10	Manejo de datos en Python: estructuras de datos, archivos.	Ciencia de datos
01/11	Librerías para manipulación de datos: NumPy. Pandas	Ciencia de datos
08/11	Limpieza, normalización y transformación de datos con Python.	Ciencia de datos
15/11	Exploración y visualización de datos con Matplotlib, Seaborn.	Ciencia de datos
22/11	Introducción a Tableau, QlikSense, Splunk, QlikView.	Ciencia de datos
29/11	Introducción a Power BI	Ciencia de datos
06/12	Técnicas de pre procesamiento de datos.	Ciencia de datos
13/12	Introducción a técnicas de aprendizaje supervisado y no supervisado con scikit learn.	Ciencia de datos
	Proyecto de ciencia de datos con Python	Proyecto final

Evaluación y acreditación

Para aprobar cada módulo, se deberá cumplimentar con los siguientes requisitos:

- Calificación del 100% de las actividades de seguimiento en el campus (como cuestionarios, participación en los foros, entre otros) y calificación de Aprobado en las actividades propuestas por cada docente durante el desarrollo del módulo.
- Presentación y defensa de un proyecto integrador del módulo.