Universidad Nacional de General Sarmiento

Los Polvorines, 09/11/2022

Resolución Rectoral N° 26874 / 2022

Aprobar el Curso de Formación Continua “Introducción a la ciencia de datos con Python”.

Expediente Digital N°5695/22.-

VISTO: el Estatuto de la Universidad Nacional de General Sarmiento; las Resoluciones (CS) N°2614/08, (CIDEI) N°6798/22; el Expediente Digital N°5695/22; y

CONSIDERANDO:

Que por Resolución (CS) N°2614/08 se aprueba los Lineamientos Generales para el Desarrollo de las actividades de Formación Continua en la Universidad Nacional de General Sarmiento.

Que mediante la resolución del Consejo del Instituto de Industria (CIDEI) N°6798/22, se aprueba el Curso de Formación Continua “Introducción a la ciencia de datos con Python”.

Que los contenidos de la propuesta se ajustan a los Lineamientos Generales para el Desarrollo de las actividades de Formación Continua de la UNGS.

Que corresponde aprobar el Curso de Formación Continua “Introducción a la ciencia de datos con Python” su estructura, contenidos y modalidad.

POR ELLO:

LA RECTORA DE LA UNIVERSIDAD NACIONAL
DE GENERAL SARMIENTO
RESUELVE:

ARTÍCULO 1°: Aprobar el Curso de Formación Continua "Introducción a la ciencia de datos con Python" cuyo programa analítico forma parte de la presente Resolución como Anexo en cuatro (4) hojas.

ARTICULO 2°: Regístrese, comuníquese al Instituto de Industria, a la Dirección de Formación Continua, a la Dirección de Títulos y Planes, a Dirección General de Biblioteca y Documentación y a la Unidad de Auditoría Interna. Cumplido, archívese.

Dra. Flavia Terigi
Rectora
Universidad Nacional de General Sarmiento

Archivos adjuntados

Nombre del archivo
Anexo I - ED5695.pdf

ANEXO I – Expediente Digital N°5695/22.-

PRESENTACION DE CURSOS – SEMINARIOS – JORNADAS

1- DATOS IDENTIFICATORIOS DEL PROYECTO

Título del proyecto: Introducción a la ciencia de datos con Python

Localización de la oferta: Dependencias de la UNGS y/o de otras instituciones con convenios con la UNGS.

Área y/o disciplina comprendida en la capacitación: Economía

Destinatarios/as: Estudiantes avanzados/as y Graduados/as de las carreras de Economía, Economía Política, Economía Industrial, Profesorado de Economía y otras carreras en ciencias sociales de la UNGS y/o de otras universidades. Investigadores/as Docentes y No Docentes de la UNGS.

Requisitos de Ingreso: Haber cursado y aprobado una materia de Estadística introductoria. Deseables: conocimientos de Econometría y lenguajes de programación.

Modalidad de la capacitación: virtual

Formato: Curso

Carga horaria: 28 horas

Costo: El curso será gratuito para los/as estudiantes de la UNGS y el personal docente y no docente de la UNGS y arancelado para los/as estudiantes externos/as.

Presupuesto: Instituto de Industria

2- EQUIPO CAPACITADOR

- **Apellido y Nombres:** Mateo Ezequiel Suster
- **N° de documento:** 38.591.901
- **Título:** Lic. en Economía Política UNGS

- **Apellido y Nombres:** Marcelo Fernando Molina
- **N° de documento:** 32.470.383
- **Título:** Lic. en Economía Política UNGS

3- DESCRIPCIÓN PROYECTO
Síntesis del proyecto:

El curso consiste en brindar una introducción al lenguaje de programación Python como herramienta para el procesamiento de datos y, en específico, para la ciencia de datos. Los temas que se abordarán serán amplios, desde

su sintaxis básica, pasando por la configuración de una base de datos, hasta el modelado de datos con técnicas de aprendizaje automático.

Fundamentación:

El desarrollo de la cuarta revolución industrial generó múltiples transformaciones en los procesos de trabajo. En especial, el aumento en la capacidad de cómputo, almacenamiento y transmisión de datos. Todos estos cambios dieron lugar al fenómeno conocido como Big Data. La multiplicidad y cantidad de datos abrió la puerta a nuevas oportunidades de generar valor utilizando los mismos como insumos. Pero, ¿qué técnicas disponemos para lograrlo?
Hoy en día, poseer herramientas de programación es un requisito fundamental para insertarse en la vida profesional. Frente a la gran variedad de lenguajes de programación, parecería difícil determinar cuál aprender para trabajar con datos. No obstante, Python es el lenguaje de programación más adecuado para dichos fines. Entre otros motivos, según el índice TIOBE, Python es hoy en día el lenguaje más popular del mundo. Esto potencia la creación de comunidades *open source*, lo cual retroalimenta la capacidad de dicha herramienta para la resolución de problemas.
Sus usos son diversos: un ejemplo es la automatización de operaciones y la construcción de aplicaciones fiables y escalables (por ejemplo, un algoritmo de ingesta de grandes cantidades de transacciones bancarias hacia un data warehouse o servidor). Hoy en día, las aplicaciones de *Machine Learning* (aprendizaje automático) se usan en una multiplicidad de industrias, tanto con fines predictivos como para encontrar patrones no explícitos en los datos.

Objetivo General:

Brindar una introducción al lenguaje de programación Python como herramienta específica de la ciencia de datos.

Objetivos específicos:

- presentar una introducción general a la ciencia de datos y de las potencialidades de Python para su estudio;
- transmitir conocimientos de manipulación y preprocesamiento de datos;
- impulsar habilidades en visualización de datos;
- generar nociones básicas para la aplicación de técnicas de aprendizaje automático.

Contenidos:

El curso se organiza en los siguientes 4 módulos:
- Módulo 1. *Introducción*
- Módulo 2. *Manipulación de datos*
- Módulo 3. *Visualización de datos*
- Módulo 4. *Introducción al aprendizaje automático (machine learning)*

Contenidos de los módulos:
Módulo 1. Introducción

1. ¿Qué es la ciencia de datos?
. La cuarta revolución industrial y el origen del *Big Data*.
a. El camino desde la estadística hacia la computación: ciencia de datos como disciplina.
b. Machine Learning: ¿es sólo estadística para el big data?

2. Elementos básicos de programación
. ¿Qué es Python?
a. Tipos de datos
b. Manipulación de listas
c. Estructuras de control
d. Funciones
e. Módulos

Módulo 2. Manipulación de datos

3. Base de datos
. Pipeline del procesamiento de datos
a. Diferencias entre Data Lake y Data Warehouse
b. Base de datos, DBMS y motores de base de datos
c. Tipos de bases de datos
d. Modelo relacional
e. SQL. Sintaxis básica
f. Ejecución de queries con SQLite

4. Dataframes
. Creación y carga de datos con Pandas
a. Indexing
b. Pivot table and join
c. Manipulación de variables. Agrupación y agregación
d. Series de tiempo: presentación y tratamiento

5. Preprocesamiento y análisis exploratorio de datos
. Limpieza, integración y preparación de los datos
a. Análisis estadístico descriptivo
b. Manipulación de datos faltantes
c. Detección de outliers
d. Feature engineering

6. APIs
. Introducción a las APIs
a. Protocolos básicos
b. Casos prácticos de APIs con datos de series de tiempo (datos.gob.ar, World Bank, FRED, etc.)

Módulo 3. Visualización de datos

7. Visualización de datos
. Cómo hacer una visualización efectiva. Ejemplos históricos
a. Nociones de graficación (forma, color, tamaño, color).
b. Tipos de gráficos
c. Creación de gráficos con Pandas
d. Utilización de otras librerías (Matplotlib, Seaborn y Plotly)

Módulo 4. Introducción al aprendizaje automático

8. Introducción al modelado estadístico
. Modelos supervisados básicos: regresión lineal.
a. Estimación, interpretación y validación (medidas de ajuste, pruebas de hipótesis, diagnósticos)
b. Trade-off sesgo varianza y sobreajuste (overfitting)
c. Medición de errores en train y testing
d. Regresión con regularización

9. Modelos de Machine Learning (supervisado)
. Tipos de aprendizaje
a. El ciclo de vida de un modelo de machine learning
b. Tuneo de hiperparámetros
c. Técnicas de validación cruzada
d. Clasificadores basados en árboles de decisión
e. Algoritmos de ensamble

10. Aprendizaje no supervisado
. Reducción de la dimensionalidad
a. Clustering

Duración y carga horaria: 28 horas totales distribuidas en 8 encuentros sincrónicos de 3 horas cada uno y un encuentro final de 4 horas.

Evaluación: Se considerará la participación en las clases de los/as estudiantes y la realización de un trabajo integrador final.

Bibliografía:

Han, J., & Kamber, M. (2006). Data mining: concepts and techniques, 2nd. University of Illinois at Urbana Champaign: Morgan Kaufmann.
Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.
González Duque, R. (2011). Python para todos.
Gowrishankar, S., Veena, A., Introduction to Python Programming, CRC, 2019.
Stephenson, B., The Python Workbook: A Brief Introduction with Exercises and
Solutions, Springer, 2019.

Certificación: Se entregarán certificados de aprobación del curso a los/las estudiantes que hayan asistido al 75% del total de la cursada y aprobado el trabajo integrador final.

Firmas

UNGS: SUDOCU
Firmado por: UNGS Sudocu
Fecha: 09/11/2022 11:09:36
Razon: Cargado por SIU-Documentos

UNGS: SUDOCU
Firmado por: UNGS Sudocu
Fecha: 09/11/2022 11:52:12
Razon: Autorizado por Flavia Terigi