

Capacitación Avanzada en Ciencia de Datos e Inteligencia Artificial

Introducción

Actualmente, existe una amplia variedad de productos y herramientas que facilitan la implementación de sistemas basados en Ciencia de Datos e Inteligencia Artificial (CD e IA), muchos de ellos de código abierto y disponibles de manera gratuita. Estas herramientas están presentes en la vida diaria como parte de las actividades y decisiones de las personas, desde el uso de asistentes virtuales, ejecución de trámites mediante formularios digitales, exploración de mapas digitales, sugerencias de productos y películas, entre otros. Sin embargo, en la mayoría de los casos no triviales, el desarrollo y mantenimiento efectivo de este tipo de sistemas requiere conocimiento y capacidades muy específicas que son difíciles de encontrar en el mercado laboral actual. La falta de estos conocimientos dificulta la formación de equipos de trabajo que puedan poner en producción sistemas basados en CD e IA. A su vez, la implementación de sistemas de CD e IA realizada sin el conocimiento sobre sus fundamentos puede resultar en riesgos importantes tanto para los responsables del desarrollo como para los usuarios finales o indirectos.

Por otro lado, la falta de capacidades y conocimiento específico puede enlentecer o detener el desarrollo de empresas proveedoras de soluciones tecnológicas, sobre todo aquellas pequeñas o medianas, impidiéndoles abordar proyectos de alto valor agregado o escalar sus soluciones de innovación en sincronía con la expansión del mercado.

Descripción del curso

Este curso tiene el objetivo de brindarte los conocimientos y habilidades técnicas, esenciales y necesarias, tanto para la implementación de las Ciencia de Datos e Inteligencia Artificial, como para conocer las implicancias de su uso. El mismo está diseñado para una audiencia diversa con conocimientos de programación.

El curso se organiza en módulos para fortalecer tu aprendizaje. En cada uno de ellos, encontrarás actividades, recursos valiosos y acompañamiento docente para comprender los fundamentos teóricos y alcanzar conocimientos prácticos para aplicar en tus esfuerzos profesionales.

En estos momentos, identificar y aplicar enfoques éticos en Ciencias de Datos e Inteligencia Artificial es crucial, por lo que este curso te capacitará para tomar decisiones informadas, reconocer estándares éticos y contribuir positivamente a tu desarrollo profesional responsable, con tu organización y con la sociedad.

Objetivos

- Proveer una formación técnica específica para personas que trabajan en empresas del sector de software que les permita capacitarse para abordar proyectos de alta complejidad.
- Promover y generar actividades de transferencia efectiva de conocimiento específico y de vanguardia desde el sistema académico y científico hacia el sector privado que desarrolla tecnología de CD e IA.
- Generar lazos y promover la interacción entre el sector científico académico y las industrias del sector tecnológico.

Por qué elegir este curso

Con este curso podrás:

- Adquirir conocimientos técnicos específicos para abordar proyectos de alta complejidad en Ciencias de Datos e Inteligencia Artificial.
- Identificar las necesidades de una Ciencias de Datos e Inteligencia Artificial responsable desde un enfoque ético y de regulaciones.

Requisitos

Poseer conocimientos avanzados de programación en algún lenguaje de programación actual (eg, Java, Python) y SQL. La capacitación se dictará en lenguaje Python.

Organización

Esta propuesta comprende un módulo troncal obligatorio de 16 semanas de duración, con 4 horas de clase por semana y luego siete módulos optativos* de 4 semanas de duración cada uno, que se dictarán en 2 horas de clase semanales. Estos módulos optativos podrán ser combinables según tu interés, incluso podés hacer todos, pero se requerirá la aprobación del módulo troncal más al menos tres optativos para acceder a la

certificación de aprobación del curso. La temática de los módulos propuestos es la que se detalla a continuación:

Además, se brindarán los siguientes módulos específicos para profundizar sobre temas de avanzada:

- Sistemas de recomendación
- Procesamiento de imágenes con redes neuronales
- Procesamiento de lenguaje natural
- Procesamiento del habla
- Procesamiento de imágenes satelitales
- Procesamiento de información geográfica
- ML Ops

Horarios de cursada

El módulo troncal se dictará los días miércoles y viernes en el horario de 9 a 11 hs., comenzando el miércoles 20 de marzo finalizando el viernes 5 de julio de 2024.

Luego del módulo troncal se contempla un receso invernal de tres semanas en el mes de julio (del 8 al 28) para acompañar a las personas que tengan personas a su cuidado.

Los módulos específicos se dictarán durante los meses de agosto, septiembre y octubre en días a confirmar, en el horario de 9 a 11.

Metodología de evaluación

Para evaluar el aprendizaje, durante el módulo troncal, se utilizará la metodología de un trabajo práctico evolutivo. Se partirá de un set de datos sobre el cual se irán aplicando los diversos conocimientos adquiridos durante el cursado y habrá distintas instancias de entrega. Esto se complementará con cuestionarios de tipo múltiple opción. Los cursos específicos, dada su duración más acotada, serán evaluados a través de un solo práctico y se complementará con cuestionarios de múltiple opción.

Condiciones para la aprobación del curso

1. 80% de asistencia a las clases sincrónicas. Se busca fomentar el aprovechamiento de las clases sincrónicas para la interacción entre docentes y alumnos/as. Eventualmente, en caso de no poder asistir se deberá enviar una justificación vía mail a pcd@fundacionsadosky.org.ar y se tendrá acceso a la clase grabada.
2. Aprobación de los exámenes.
3. Aprobación de los trabajos prácticos.

Certificación

Las personas que aprueben el curso recibirán el correspondiente certificado de aprobación consignando los módulos aprobados y la cantidad de horas.

Modalidad

Virtual sincrónica, a través de la plataforma zoom. Las clases serán grabadas y podrán ser accedidas a pedido.

Cupo

30 personas

Carga horaria semanal por módulos

Módulo troncal obligatorio: 4 horas semanales

Módulos optativos: 2 horas semanales

Capacitadores

Módulo troncal: 2 docentes

Módulos optativos: 1 docente

Duración

8 meses

Fecha de inscripción

Desde el 19 febrero al 8 marzo de 2024

Fecha de confirmación de cupo

12 de marzo

Costo de inscripción

\$638.000

Si trabajás en una empresa de Argentina, sos estudiante o docente de una universidad pública o trabajás en un organismo público, tenés un 10% de descuento en tu inscripción.

Precio con 10% de descuento

\$574.200

Modalidad de pago

Por transferencia bancaria en:

- 1 pago de \$638.000
- 3 cuotas de \$245.000 (a pagar del 1 al 10 de marzo, abril y mayo)

Con 10% de descuento

Por transferencia bancaria en:

- 1 pago de \$574.200
- 3 cuotas de \$220.110 (a pagar del 1 al 10 de marzo, abril y mayo)

Cómo inscribirse

A través de este [Formulario de inscripción](#)

Consultas

pcd@fundacionsadosky.org.ar

Programa

Ver [Anexo I](#)

Anexo I: Programa

Marzo - Julio 2024

Módulo 1: Análisis de datos y aprendizaje automático

1.1. Análisis exploratorio de datos

Bibliotecas: Pandas y Seaborn.

Estadística descriptiva. Medidas de resumen: media, mediana, cuartiles, percentiles, desvío estándar.

Preprocesamiento y preparación de datos numéricos. Estandarización, normalización, detección y remoción de outliers, datos faltantes. Limpieza de datos.

Distribuciones de probabilidad.

Visualización de datos. Tipos de plots básicos y cuándo usarlos: scatter-plots, bar-plots, box-plots, violin-plots.

1.2. Modelos de regresión

Biblioteca: scikit-learn.

Introducción a scikit-learn.

Regresión lineal, cuadrados mínimos. Funciones de costo. Regresión polinomial.

Concepto de sobreajuste. Train-test, cross validation.

Regresión logística. Clasificación vs. regresión.

Optimización, método de descenso del gradiente y otros.

1.3. Modelos de clasificación

Métricas para la medición de performance. Matriz de confusión.

Clasificador lineal.

Clasificación basada en instancias: KNN.

Árboles de decisión y Random forest. Concepto de Ensembles.

Support vector machines. Kernels.

1.4. Datos faltantes

Caracterización, estrategias utilizadas para su tratamiento.

Interpolación y técnicas básicas de imputación.

Correlación, lags y otras aventuras. Riesgos y potenciales consecuencias.

1.5. Visualización de datos

Tipos de datos. Tipos de variables. Gráficos fundamentales. Herramientas para la elección de paletas de colores.

La visualización de datos. Los datos cuentan historias. Adaptación a la audiencia.

Formulación de preguntas a responder con los datos. Transformación de datos. Marcas y canales: expresividad y efectividad.

1.6. Aprendizaje no supervisado, reducción de la dimensión

Clustering de datos. K-means.

La maldición de la dimensión.

Estrategias de selección de variables: basadas en filtros, eliminación recursiva (BIC, akaike).

Reducción de la dimensionalidad. PCA.

Otros métodos de clustering y reducción de la dimensionalidad. t-distributed Stochastic Neighbor Embedding (t-SNE).

1.7. Introducción a las redes neuronales

Introducción. Perceptrón. Perceptrones multicapa.

Funciones de activación.

Grafos. Capa densa.

1.8. Funciones de pérdida

Distintas funciones de pérdida.

Técnicas de regularización y de optimización.

Aplicación del descenso por gradiente, descenso por gradiente estocástico, otras técnicas.

Funciones convexas y no convexas.

1.9. Redes neuronales

Descenso por gradiente. Backpropagation. Tensores.

Construcción de modelos.

Lectura y escritura de modelos en disco.

1.10. Transfer learning

Redes pre-entrenadas, Fine-tuning, Aplicaciones.

1.11. Ética de la IA en la práctica

Ética en Ciencia de Datos. Sesgos. Casos donde los sesgos de los modelos de IA han resultado dañinos.

Inspección de los datos. Responsabilidad compartida en la creación de datasets.

Análisis de casos prácticos. Herramientas disponibles para el diagnóstico. Estrategias de mitigación

Agosto 2024

Módulo 2: Sistemas de recomendación

Introducción. Tareas de recomendación.
Filtros colaborativos. Retroalimentación implícita y explícita.
Algoritmo de factorización matricial. Modelos con autocodificadores.
Ranking personalizado con enfoque Bayesiano.

Agosto 2024

Módulo 3: Procesamiento de imágenes con redes neuronales

Introducción al procesamiento de señales, adquisición, representación, filtrado. Convolución. Aplicaciones.
Redes neuronales convolucionales (CNN)
Principales arquitecturas para la clasificación de imágenes.
Introducción a la derivación automática. Framework PyTorch. Práctica con Python Notebooks en la nube (Colab).
Redes neuronales convolucionales para la segmentación de imágenes. Arquitecturas clásicas para segmentación. Redes totalmente convolucionales. Autocodificadores.

Septiembre 2024

Módulo 4: Procesamiento del lenguaje natural

Introducción al procesamiento del lenguaje natural (NLP). Redes neuronales recurrentes (RNN). Embeddings de palabras: Word2Vec, GloVe, fastText.
Dinámica de las redes haciendo un estudio sobre las matrices de recurrencia.
Fundamentos, topologías y algunas aplicaciones en Machine Learning.

Septiembre 2024

Módulo 5: Procesamiento del habla

Elementos de acústica y procesamiento digital de señales. Elementos de fonética, fonología y prosodia.
Síntesis del habla: concatenativa, formantes, articulatoria y modelos generativos.
Sistemas de texto-a-habla (TTS): front-end y back-end.
Reconocimiento del habla. Modelos ocultos de Markov (HMM) y redes neuronales profundas.

Sistemas de diálogo. Comprensión y generación del lenguaje hablado.
Evaluación de sistemas de procesamiento del habla.

Octubre 2024

Módulo 6: Procesamiento de imágenes satelitales

Sensores remotos ópticos y de radar. Firmas espectrales.
Introducción al uso de Python para el procesamiento de imágenes satelitales.
Imágenes georeferenciadas. Librería GDAL. Operaciones sobre archivos raster espaciales.
Índices. Datos vectoriales en Python. Librería Geopandas. Librería Rasterio.
Análisis exploratorio de datos: El espacio de atributos.
Aprendizaje automático: Métodos supervisados y no supervisados. Clasificadores supervisados con librería Scikit-Learn. Métricas de error de clasificaciones supervisadas. Clasificadores no supervisados en Scikit-Learn. Selección de modelos.

Octubre 2024

Módulo 7: Procesamiento de Información Geográfica

Procesamiento de datos vectoriales con la librería de Python Geopandas.
Fundamentos teóricos para el manejo de información geográfica (proyecciones y sistemas de referencia), Shapely y los distintos tipos de objetos geométricos (líneas, puntos y polígonos).
Manipulaciones geométricas: centroides, distancias, coordenadas, áreas, etc. Joins espaciales.
Mapas y recursos de visualización: Leaflet, Folium, Contextly, Matplotlib, etc.

Octubre 2024

Módulo 8: ML Ops

Ciclo de vida de un modelo de Machine Learning. Componentes de ML Ops
Enfoques para la implementación de modelos de ML. Despliegue de modelos en diferentes entornos (local, cloud, edge). Monitoreo y medición del rendimiento del modelo.
Herramientas de automatización
Configuración de entornos de desarrollo y producción. Manejo de dependencias y versiones de paquetes. Herramientas para el control de versiones de modelos. Integración continua y entrega continua
Gestión de datos: Construcción de pipelines de recolección, limpieza, etiquetado, validación, procesamiento y transformación de los datos.
Implementación de métricas de fairness. Explicabilidad de los modelos

Seguridad y cumplimiento normativo. Herramientas y técnicas para garantizarlas.