

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.

Para responder a esta necesidad, desde el año 2023 la Fundación Sadosky ofrece esta Capacitación Avanzada en CD e IA, dirigida a profesionales que tengan conocimientos de programación y deseen iniciarse en el camino de la ciencia de datos y la IA.

A lo largo de esta capacitación, los participantes adquirirán una base sólida en conceptos, metodologías y herramientas de CD e IA, preparándose para enfrentar los desafíos de su implementación y comprendiendo la importancia del uso responsable de datos y algoritmos en distintos ámbitos.

Descripción de la capacitación

Esta capacitación tiene el objetivo de brindarte los conocimientos y habilidades técnicas, esenciales y necesarias, tanto para la implementación de herramientas de Ciencia

de Datos e Inteligencia Artificial, como para conocer las implicancias de su uso. La misma está diseñada para una audiencia diversa con conocimientos de programación.

En estos momentos, identificar y aplicar enfoques éticos en Ciencia 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 esta capacitación

Con esta capacitación podrás:

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

Programa

Abril - Agosto

Curso 1: Análisis de datos y aprendizaje automático

1.1. Fundamentos para empezar en CD e IA

Conceptos y metodología.
Herramientas, instalaciones y/o usos.
Reforzando programación en Python.

1.2. Análisis exploratorio de datos

Fundamentos del Análisis exploratorio de datos (EDA)
Manipulación de datos: Pandas.
Estadística descriptiva. Medidas de resumen.
Preprocesamiento y preparación de datos. Estandarización, normalización, detección y remoción de outliers.
Distribuciones de probabilidad: distribución uniforme y normal.

1.3. Datos faltantes

Caracterización. Tipos.
Detección de valores faltantes: Técnicas y visualización.
Métodos de imputación: Básicas y avanzadas.

1.4. Visualización de datos

Fundamentos. Tipos de datos. Tipos de variables.
Herramientas de visualización: Matplotlib y Seaborn.
Narrativa con datos.

1.5. Modelos de clasificación

Introducción a scikit-learn.
Clasificador lineal.
Clasificación basada en instancias: KNN.
Árboles de decisión.
Support vector machines.
Métricas para la medición de performance. Matriz de confusión.
Concepto de sobreajuste. Train-test, cross validation.

1.6. Modelos de regresión

Regresión lineal, cuadrados mínimos. Funciones de costo. Regresión polinomial.
Métricas para la medición de performance.
Regresión logística. Clasificación vs. regresión.
Optimización, método de descenso del gradiente y otros.
Concepto de sobreajuste. Train-test, cross validation.

1.7. Aprendizaje no supervisado, reducción de la dimensión

Clustering de datos. K-means.
La maldición de la dimensión.
Reducción de la dimensionalidad. PCA.

1.8. Introducción a las redes neuronales

Introducción. Perceptrón. Perceptrones multicapa.
Funciones de activación.
Grafos. Capa densa.
Función de pérdidas.

Agosto

Curso 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.

Curso 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

Curso 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.

Curso 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.

Curso 6: 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.

Octubre

Curso 7: 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.

Curso 8: 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.