

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

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. 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.

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.

Requisitos

Poseer conocimientos avanzados de programación en algún lenguaje de programación actual (eg, Java, Python) y SQL. La capacitación se dictará en lenguaje Python. Se priorizará a quienes trabajen en empresas del sector software.

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

Organización

Esta propuesta comprende un módulo troncal obligatorio de 14 semanas de duración, con 5 horas de clase por semana (módulo 1) y luego siete módulos optativos* de 4 semanas de duración cada uno que se dictarán en 2 horas de clase semanales (módulos 2 al 8). Estos módulos optativos podrán ser combinables según tu interés, incluso podés hacer todos, pero se requerirá la aprobación del módulo troncal más al menos dos optativos para acceder a la certificación de aprobación del curso. La temática de los módulos propuestos es la que se detalla a continuación:

Módulo 1: Análisis de datos y aprendizaje automático

Módulo 2: Sistemas de recomendación

Módulo 3: Procesamiento de imágenes con redes neuronales

Módulo 4: Procesamiento del lenguaje natural

Módulo 5: Procesamiento del habla

Módulo 6: Procesamiento de imágenes satelitales

Módulo 7: Procesamiento de información geográfica

Módulo 8: ML Ops

*El dictado de cada módulo optativo queda sujeto a que se cubra el cupo mínimo de 5 personas inscriptas.

Docentes



Matías Agustín Pérez

MINI BIO

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

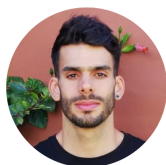
Me llamo Matías Agustín Pérez y tengo 28 años. Soy ingeniero en Sistemas, graduado de la Universidad Nacional de Chilecito, donde adquirí una sólida formación en el campo de la tecnología y la programación. A lo largo de mi carrera, me he destacado como docente investigador, con una amplia experiencia en la enseñanza de asignaturas como Inteligencia Artificial, Programación, Estructura de Datos y Cálculo Numérico.

Mi pasión por la educación me ha llevado a dedicarme a la docencia en varios niveles durante los últimos 10 años. He tenido el privilegio de impartir conocimientos y compartir mi entusiasmo por la tecnología en diferentes programas y proyectos. Entre ellos se encuentran cursos de Full Stack en el plan de Argentina Programa, 111K Programadores, y La Programación y su Didáctica Método Program.AR de la Fundación Manuel Sadosky.

Además de mi trabajo como docente, también estoy enfocado en continuar mi desarrollo profesional. Actualmente, me encuentro finalizando mi especialización en Inteligencia de Datos Orientada a Big Data, donde he estado inmerso en el estudio y aplicación de las últimas tecnologías y metodologías relacionadas con el manejo y análisis de grandes volúmenes de datos.

Para mí, la educación es un proceso continuo y bidireccional. Disfruto transmitir mis conocimientos a los estudiantes, pero también valoro enormemente la oportunidad de aprender de ellos. La interacción y el intercambio de ideas son fundamentales para enriquecer el proceso educativo.

En resumen, mi trayectoria como ingeniero en Sistemas y docente investigador me ha permitido adquirir una sólida formación en diversas áreas de la tecnología y la programación. Estoy emocionado por formar parte de este curso y seguir creciendo tanto profesional como personalmente. Estoy comprometido a compartir mi experiencia y conocimientos, así como a aprender de cada uno de ustedes durante este emocionante viaje de aprendizaje.



Martín Rodríguez Núñez

MINI BIO

¡Hola! Soy Martín Rodríguez Núñez, un ingeniero ambiental y científico de datos apasionado por aplicar la inteligencia artificial en resolver problemas del mundo real. Actualmente, estoy inmerso en un doctorado en ciencias de la ingeniería, enfocado en la aplicación de la ciencia de datos e inteligencia artificial en el análisis de datos de contaminantes atmosféricos. Paralelamente, me

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

encuentro en el tramo final de mi maestría en estadística aplicada, donde defenderé mi tesis en los próximos meses.

Durante mi carrera profesional, he participado en múltiples proyectos que involucran la ciencia de datos e inteligencia artificial, lo que me ha permitido adquirir habilidades prácticas y conocimientos especializados en el área. Además, he tomado numerosos cursos de posgrado y he asistido a conferencias para mantenerme actualizado con las últimas tendencias y técnicas de la disciplina. Me apasiona y estoy entusiasmado de poder compartir mis conocimientos y experiencia en este curso. Fuera del ámbito laboral, disfruto de la naturaleza y el deporte. Espero poder inspirar y motivarlos a lo largo del curso para que puedan llevar la ciencia de datos e inteligencia artificial a nuevos horizontes. ¡Espero verlos pronto!



Enzo Ferrante

MINI BIO

Enzo estudió Ingeniería de Sistemas en la UNICEN (Tandil, Argentina), realizó el Doctorado en Informática en la Université Paris-Saclay y el INRIA (París, Francia), fue alumno visitante en la Universidad de Stanford (California, EEUU) e investigador Fulbright en el Massachussets General Hospital y la Escuela de Medicina de Harvard (Boston, EEUU), realizó su postdoctorado en el Imperial College London (Londres, Reino Unido) y a fines de 2017 volvió a la Argentina como investigador repatriado al sinc(i): Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional dependiente del CONICET y la Universidad Nacional del Litoral. Allí lidera la línea de investigación en Machine Learning para el análisis de imágenes biomédicas. Es también docente en la UNL, en la Universidad Torcuato Di Tella y en la Universidad de San Andrés. En el año 2020 recibió el premio Estímulo de la Academia Nacional de Ciencias Exactas, Físicas y Naturales, y el premio Ciencia y Tecnología del Mercosur, por sus contribuciones al área de inteligencia artificial y el análisis de imágenes biomédicas. Es además miembro de la Academia Joven de Argentina.

Horarios de cursada

El módulo troncal se dictará los días miércoles y viernes en el horario de 9 a 11:30 hs., comenzando el

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

miércoles 7 de junio y finalizando el viernes 29 de septiembre de 2023. Se contempla un receso invernal de tres semanas en el mes de julio (del 10 al 28) para acompañar a las personas que tengan hijos a cargo.

Los módulos específicos se dictarán en horarios a confirmar, pero estarán siempre dentro de la franja horaria de 9 a 18 hs.

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. 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.

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

Cupo

30 personas

Carga horaria semanal por módulos

Módulo troncal obligatorio: 5 horas semanales

Módulos optativos: 2 horas semanales

Capacitadores

Módulo troncal: 2 docentes

Módulos optativos: 1 docente

Duración

6 meses

Fecha de inscripción

Del 23 al 30 de mayo

Fecha de confirmación de cupo

2 de junio

Costo de inscripción

\$220.500

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

Modalidad de pago

Por transferencia bancaria en:

- 1 pago de \$220.500
- 3 cuotas de \$84.525 (a pagar en junio, julio y agosto)

Cómo inscribirse

[Formulario de inscripción](#)

Consultas

pcd@fundacionsadosky.org.ar

Programa

Ver [Anexo I](#)

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

Anexo I: Programa

Junio - Septiembre 2023

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.

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

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.

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

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

Octubre 2023

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

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.

Octubre 2023

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.

Octubre 2023

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.

Noviembre 2023

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

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.

Noviembre 2023

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.

Noviembre 2023

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.

CAPACITACIÓN EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

ORIENTADA A PERSONAS QUE TRABAJAN EN LA INDUSTRIA DEL SOFTWARE

Mapas y recursos de visualización: Leaflet, Folium, Contextly, Matplotlib, etc.

Noviembre 2023

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.