

**SISTEMA NACIONAL DE CAPACITACION
DISEÑO DE LA ACTIVIDAD****Nombre**

PROGRAMA INICIAL EN CIENCIA DE DATOS: INTRODUCCIÓN AL APRENDIZAJE AUTOMÁTICO

Código INAP IN37394/22 **Estado** Activo**Programa**)Campos de Práctica **Área** Sistemas, procesos y tecnologías**Fundamentación**

Propósito: Desarrollo o fortalecimiento de capacidades.

En la actualidad se produce a nivel global una creciente cantidad de datos que se constituyen en un insumo estratégico para la construcción de información relevante para la toma de decisiones en las organizaciones y para la producción de más y mejores resultados y, como en el ámbito del sector público resulta clave fortalecer las capacidades de las y los trabajadores para el tratamiento de los datos que se producen, la Secretaría de Gestión y Empleo Público y el Instituto Nacional de Estadística y Censos (INDEC) celebraron un convenio de cooperación técnica que, entre otras, tiene por finalidad desarrollar, promover y ejecutar proyectos de capacitación en temas vinculados con el tratamiento de datos.

En esta convergencia de valoración, el INDEC ha organizado el PROGRAMA INICIAL EN CIENCIA DE DATOS cuyo objetivo es desarrollar competencias para la búsqueda, proceso, interpretación y análisis de los resultados. Por su parte, el Instituto Nacional de la Administración Pública diseñó el Programa INAP FUTURO que tiene por finalidad conformar un espacio de reflexión, planificación, experimentación e implementación sistemática basado en el análisis de los escenarios posibles del futuro y de las transformaciones sociales y organizacionales del presente.

El curso Introducción al Aprendizaje Automático se inscribe en el marco de los programas INAP FUTURO y PROGRAMA INICIAL EN CIENCIA DE DATOS y se propone dar respuesta a las necesidades de las instituciones públicas que, atravesadas por un contexto altamente dinámico en lo tecnológico, demandan eficiencia y eficacia en la concreción de sus funciones. Por eso, se considera necesaria la incorporación de herramientas y técnicas de inteligencia artificial, para identificar y determinar patrones de comportamiento de variables en la gestión de datos: un conocimiento entendido como aprendizaje automático.

A partir de lo mencionado y en línea con la Propuesta Formativa del INAP, en la presente actividad prevalecen los siguientes tipos de saberes: Saber (saberes objetivados sobre la realidad organizados en sistemas de conceptos y teorías)
- Saber hacer (saberes de acción vinculados con la capacidad de intervenir).

Contribución esperada

Se espera que los profesionales y técnicos de la Administración pública desarrollen competencias que les permitan realizar una actualización teórico-metodológica en el procesamiento inteligente de los sistemas de información.

Perfil del participante

Agentes profesionales y técnicos que requieran actualizar conocimientos de aprendizaje automático para la gestión, análisis, e inferencia de datos.

Objetivos

Se espera que los participantes logren:

- actualizar conocimientos sobre la preparación preliminar de los datos, la exploración y el perfilado estadístico de la información;
- incorporar técnicas del aprendizaje automático para el planteo, construcción y evaluación de modelos de clasificación, predicción y segmentación;
- conocer nuevas prácticas progresivas de distintos casos de uso y aplicación.

Contenido

UNIDAD 1: MACHINE LEARNING CLASIFICACIÓN

Tipos de problemas de machine learning.

Aprendizaje supervisado vs. no supervisado.

Ciclo de trabajo de un modelo predictivo.

Armado del dataset.

Entrenamiento y testeo.

Overfitting y underfitting.

Árboles de decisión.

Regresión logística.

Curvas ROC.

Evaluación de resultados y casos de uso.

UNIDAD 2: MACHINE LEARNING REGRESIÓN LINEAL

Fundamentos de la regresión lineal.

Supuestos y estimación de parámetros.

Selección de variables.

Adecuación.

Reducción de la dimensionalidad.

Validación.

Predicción.

Casos de uso.

UNIDAD 3: MACHINE LEARNING CLUSTERING

Definición de cluster.

Aplicación en el campo de datamining.

Evaluación de similitudes.

Medidas de distancia.

Clustering jerárquico.

K-Means.

PAM.

DBSCAN.

Fuzzy Clustering.

Evaluación de clusters.

Validación de resultados.

El problema de la dimensionalidad.

UNIDAD 4: MACHINE LEARNING TOOLBOX

Pipelines de Scikit-Learn.

Paquete Caret de R.

Selección óptima de parámetros.

Optimización de la función objetivo.

Estrategias y principios de ensamble.

Bootstrapping.

Bagging.

Boosting.

Random Forest.

XGBoost.

Ridge.

Lasso.

Estrategias metodológicas y recursos didácticos

Las clases serán de modalidad será teórico-práctica con actividades sincrónicas y asincrónicas. Se desarrolla en el Campus Virtual del INDEC.

Las clases sincrónicas se podrán iniciar con una revisión de los contenidos estudiados en la clase anterior o con la aclaración o ampliación de las devoluciones de las actividades asincrónicas; se incentivará la participación de los estudiantes. Se continuará con el desarrollo de contenido planificado para la clase y se cerrará con la integración del

contenido abordado en la clase desarrollada, con exposición dialogada o con el planteo de ejercitaciones.

En las instancias asincrónicas, se presentarán actividades para resolver durante una semana, con fecha de entrega hasta el inicio de la próxima clase sincrónica. Durante esa semana de instancia asincrónica, se podrán realizar consultas a través de un foro destinado a tal fin y recibir, por parte del docente, las devoluciones correspondientes que inviten a la reflexión y sirvan de instrumento para la construcción del conocimiento.

Se alternarán las siguientes metodologías didáctico-pedagógicas:

- Exposición participativa con presentación de diapositivas y videos.
- Trabajo en equipos.
- Presentación de casos y ejemplos.
- Trabajos prácticos.
- Devoluciones escritas sobre las actividades asincrónicas.

Recursos didácticos:

- Plataforma web del INDEC
- Presentaciones de diapositivas.
- Videos.
- Casos escritos.

Descripción de la modalidad

Mixta (sincrónica y asincrónica).

Bibliografía

- Jones, H (2019): "Aprendizaje Automático". Ed. Bravex Publications. Español.
- Dinov. I. (2020): "Data Science and Predictive Analytics". Ed. Springer International Publishing AG. Inglés.
- Jes & Uacute y Bobadilla Sancho, S. (2020): "Machine Learning y Deep Learning". Ed. Ra-ma. Español.
- Bishop. C. (2020): "Pattern Recognition and Machine Learning". Ed. Springer Verlag New York Inc. Inglés.
- Schmarzo, B. (2014): "Big data, el poder de los datos". Ed. Anaya Multimedia. Español.

Evaluación de los aprendizajes

Evaluación de proceso: Los participantes realizarán actividades de aprendizajes, de resolución de cuestionarios, análisis de casos y trabajos prácticos de resolución individual, en parejas o grupales. Estas actividades tendrán una devolución por parte del docente, la cual servirá de instrumento para la construcción del conocimiento.

Evaluación de producto: Los participantes realizarán una evaluación final que contemplará los contenidos temáticos del curso; será resuelta de manera individual para ser corregida por el docente. Esta evaluación final constará, además, con un múltiple choice de 10 ítems que deberá ser completado con respuesta única. Cada ítem tendrá valor de un punto.

Se aprobará la evaluación según los siguientes criterios:

- Cuando la respuesta sea precisa y se acompañe con una fundamentación que dé cuenta de la aplicación de conceptos y/o herramientas de los marcos teóricos y metodológicos abordadas en el curso.
- Resolución de manera correcta del 70 % o más de los ítems del múltiple choice.

Instrumentos para la evaluación

- Grilla de corrección: para la evaluación de proceso, se construirá de acuerdo a los “aspectos positivos y a mejorar”; para la evaluación de producto, se construirá de acuerdo al contenido temático.
- Encuesta de satisfacción para la evaluación del curso. Se construirá de acuerdo a distintos aspectos pedagógico-didácticos: desempeño del/la docente, los contenidos temáticos, de desarrollo del curso. Asimismo, se solicitarán sugerencias y observaciones como propuesta de mejora. Se administrará al finalizar la cursada, de forma anónima para conocer la opinión de los participantes.

Requisitos de Asistencia y aprobación

Los asistentes deberán acreditar:

- asistencia mínima del 75% en las clases sincrónicas;
- realización y aprobación de las actividades prácticas propuestas por el docente en la modalidad asincrónica;
- aprobación del trabajo práctico integrador final con nota mínima de siete (7) en una escala de 1 a 10. Si el trabajo final no alcanza la nota para su aprobación tendrá la posibilidad de una segunda presentación.

Duración (Hs.)

24

Detalle sobre la duración

Las clases sincrónicas se desarrollarán en CUATRO (4) encuentros virtuales de TRES (3) horas de duración cada uno.
Total: DOCE (12) horas reloj.

Las instancias asincrónicas estarán destinadas a la lectura de la bibliografía y la realización de las actividades para su posterior devolución por parte del docente. Se estima una carga horaria de TRES (3) horas para la resolución por clase.

Total: DOCE (12) horas reloj.

Lugar

Campus Virtual del INDEC: campusvirtual.indec.gob.ar

Perfil Instructor

Especialista en ciencia de datos e inteligencia artificial con formación en grado y/o posgrado.

Origen de la demanda

INAP-INDEC

Prestadores Docentes

CUIT/CUIL	APELLIDO Y NOMBRE
27273398320	BERAUDO,VANINA ELISABET
20149519379	COPOLILLO,NESTOR RAUL
20138079954	POGGI,EDUARDO ALBERTO