

Propuesta Actividad de Posgrado

Resolución N° 442/21C.S.

1. Tipo de actividad (curso, seminario o taller)

Indicar si el curso es de Posgrado o de Actualización y Perfeccionamiento

Curso de postgrado

2. Denominación

Consignar el nombre de la actividad, teniendo en cuenta que sea claro y conciso.

Técnicas de aprendizaje automático – machine learning

3. Unidad Académica Responsable:

Facultad de Ciencias Exactas y Naturales y Agrimensura – UNNE

4. Fundamentación:

Referirse brevemente a la necesidad que dio origen a la propuesta, qué demanda se estaría atendiendo con su dictado, a qué rama del saber se aporta; contextualización, justificación y campo disciplinar.

Una de las áreas que ha tenido un significativo avance en los últimos años es la ciencia de datos y en particular la inteligencia artificial y el aprendizaje profundo, con un importante impacto en la sociedad en general como también en muy diversas áreas de investigación. Para que este avance fuera posible, fue necesaria la coexistencia de tres aspectos fundamentales, como ser la existencia de importantes bases de datos digitales, poder de cómputo matricial (GPU) y el desarrollo de técnicas estadísticas basadas en redes neuronales profundas que permiten el aprendizaje de diversas características de los datos en muy alta dimensionalidad. La conjunción de estos tres aspectos dio lugar a significativos avances en las más diversas aplicaciones, desde la interpretación de imágenes al procesamiento y la generación de textos. El poder de predicción o clasificación del aprendizaje estadístico abarca desde bases de datos de imágenes o textos, hasta datos de la dinámica molecular o el clima.

El dictado de este curso surge como respuesta a requerimientos planteados por la propia institución que busca la formación de los docentes, investigadores y egresados en la ciencia de datos y en el estado del arte del aprendizaje automático.

5. Objetivos:

Señalar los objetivos de aprendizaje que se persiguen con el dictado de la actividad formativa; los conocimientos, habilidades y actitudes que se pretenden que logren los participantes.

(1) Introducir los conceptos del aprendizaje automático con fundamentación estadística y con una perspectiva algorítmica. Desarrollar en el estudiante un pensamiento crítico y habilidades de razonamiento independiente en el área.

(2) Introducir las nociones básicas relacionadas con la preparación de bases de datos de distintas disciplinas (Química, Física, Matemática, Biología, Ingeniería)

para la posterior aplicación de técnicas de aprendizaje automático.

(3) Desarrollar aplicaciones de aprendizaje automático de interés para empresas tecnológicas y disciplinas y líneas de investigación de las Ciencias Exactas y Naturales.

(4) Introducir herramientas de código abierto para la implementación de técnicas de aprendizaje automático.

(5) Generar un conjunto de códigos que sirvan de base y guía para la aplicación de técnicas de aprendizaje automático en distintas áreas científicas y bases de datos.

6. Destinatarios:

Consignar la titulación de grado y los perfiles de los destinatarios de la actividad formativa

Egresados de carreras de ciencias exactas y naturales y todas las ingenierías. Se requiere de conocimientos previos de un curso de estadística básico y de conocimiento de programación (se trabajará durante el curso en lenguaje python)

7. Fecha de inicio y fecha de finalización:

15 de Agosto 2023 – 15 de Noviembre 2023

8. Modalidad:

Indicar si es presencial - con más del 50% de la carga horaria presencial - o virtual - con más del 50% de la carga horaria no presencial, a distancia

La modalidad será dual: Presencial o Virtual a definir por los estudiantes en el momento de la inscripción.

Los estudiantes de carácter Presencial tendrán el total de 60 horas todas presenciales. Los estudiantes de Carácter Virtual tendrán el total de 60 horas sincrónicas virtuales. Para los estudiantes Virtuales se utilizarán las tecnologías: google-meet, zulip y slack.

9. Carga horaria:

Consignar la carga horaria total considerando que la carga horaria mínima de los cursos de posgrado es de 30 horas y de 120 horas máximas; los talleres o seminarios pueden desarrollarse con un mínimo de 15 horas, hasta un máximo de 30 horas; si corresponde indicar el porcentaje de horas virtuales y de horas presenciales

Estudiantes Presenciales:

Carga Horaria total: __60__hs

% horas virtuales: ____

% horas presenciales: __60__

Estudiantes Virtuales

Carga Horaria total: __60__hs

% horas virtuales: __60__

% horas presenciales: ____

10. Cupo:

Especificar cupo máximo y mínimo

Cupo Máximo: 30 (cupo correspondiente a los alumnos presenciales)

Cupo Máximo: 10 (cupo correspondiente a los alumnos virtuales)

Cupo Mínimo: 10

11. Certificaciones a otorgar

(Solo se otorgará certificación de aprobación del curso, seminario o taller)

Certificado de aprobación

12. Condiciones a cumplir para la emisión del certificado

Aprobación del trabajo final asignado.

13. Contenidos:

Indicar los contenidos que se desarrollarán durante el cursado, según el criterio de organización adoptado - unidades, módulos, etc.-; la cantidad de contenido debe ser acorde a las horas de dictado, se deben incluir contenidos relevantes para una formación de posgrado.

Los contenidos están subdivididos en 12 módulos cada uno de ellos con una carga horaria de 5 horas (2 horas de teoría y 3 de prácticas).

1. Introducción general
2. Introducción a python, numpy, pytorch.
3. Estadística. Teoría de probabilidad. Regresión. Máxima verosimilitud. Maximum a posteriori estimación.
4. Modelos lineales: regresión y clasificación.
5. Árboles y bosques aleatorios
6. Clustering y reducción de la dimensionalidad.
7. Máquinas de soporte vectorial. Nucleos/kernels.
8. Redes neuronales. Entrenamiento. Backpropagation.
9. Redes convolucionales.
10. Redes recurrentes. LSTM y GRU.
11. Aprendizaje profundo. UNET.
12. Transformers. Atención. Modelos de lenguaje

14. Bibliografía básica:

Enumerar los textos básicos que serán manejados total o parcialmente durante el dictado, que den cuenta del enfoque adoptado y su actualización; en la modalidad a distancia, la bibliografía deberá estar disponible en formato digital en plataformas o repositorios accesibles

Bishop C., 2006: *Pattern Recognition and Machine Learning*. Springer.

Goodfellow I., Bengio Y., Courville A., 2017. *Deep Learning*. MIT Press

Murphy, K.P., 2012. *Machine learning: a probabilistic perspective*. MIT press.

James, G., Witten, D., Hastie, T. and Tibshirani, R., 2013. *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.

Zhang, A., Lipton, Z.C., Li, M. and Smola, A.J., 2021. *Dive into deep learning*. arXiv preprint arXiv:2106.11342.

Stevens, E., Antiga, L. and Viehmann, T., 2020. *Deep learning with PyTorch*. Manning Publications.

Géron A., 2019: *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media Inc.

15. Metodología de enseñanza:

Consignar las estrategias de enseñanza que se priorizarán en el dictado presencial, las actividades que se desarrollarán para alcanzar los objetivos propuestos; si la propuesta es a distancia, indicar: actividades individuales y grupales de producción, análisis e intercambio colaborativo utilizando las herramientas TIC; actividades que promueven el desarrollo autónomo en tareas conjuntas focalizando la investigación y búsqueda, interacción con pares en tareas colaborativas, análisis de casos, resolución de situaciones problemáticas; actividades de seguimiento y tutorías para la elaboración de informes, monografías, trabajos prácticos, actividades de diagnóstico, de reflexión, de elaboración, de intercambio y discusión, de relevamiento de datos, actividades relacionadas con la práctica profesional; empleando herramientas de diferentes soportes: confección de hipertextos, foros, wikis, videos, guías, autoevaluación, simulación o modelización; otras

Dictado de teóricos (Dos horas semanales). Desarrollo de trabajos prácticos (Tres horas semanales) en computadoras en grupos de hasta dos, utilizando herramientas Jupyter Notebooks, Google Colab, y se dispondrá de las plataformas zulip para la interacción y consultas, para el trabajo con las bases de datos se utilizará Google Drive y Gitlab para el intercambio/versionado de códigos.

16. Materiales didácticos a utilizar:

Describir los materiales de estudio básicos que se requerirán para el cursado – módulos didácticos, guías de estudio, recursos virtuales, herramientas de simulación u otros.)

Guías de prácticos basadas en jupyter notebooks. Bases de datos. Librerías de código abierto.

17. Instancias de evaluación y aprobación:

Describir detalladamente la metodología de evaluación. En el caso de propuestas a distancia indicar los instrumentos y soportes que se emplean para evaluar los aprendizajes y competencias de los alumnos incorporando formas de evaluación innovadoras –sincrónicas, asincrónicas, colaborativas, etc.- a partir de los recursos que ofrece la plataforma Moodle u otras que promuevan el uso de las TIC - Res. N°285/18 C.S. La calificación se realizará conforme a la Escala de Calificaciones vigente de la UNNE

La evaluación del curso consistirá en un trabajo final que consiste en la aplicación de alguna/s de las técnicas introducidas durante el curso para la resolución de un problema específico utilizando conjuntos de datos provistos durante el curso. En una exposición, el estudiante o en grupo de dos deberá/n exponer el desarrollo, fundamentar la técnica introducida y analizar los resultados obtenidos.

18. Modalidad de la evaluación final

(Individual o grupal).

Puede ser individual o grupal (de hasta dos) .

19. Requisitos de aprobación del curso:

Enumerar cuáles serán las exigencias para otorgar la aprobación del curso, taller o seminario, además de cumplir con las evaluaciones anteriormente mencionadas, por ejemplo, asistencia, pago de arancel, etc.).

Asistencia al 80% de las clases. Entrega del código y presentación oral del trabajo final asignado. La entrega del código y la presentación oral serán presenciales tanto para los estudiantes presenciales como para aquellos virtuales.

20. Cronograma estimativo:

Consiguar cómo se distribuirán las horas de dictado de la actividad formativa en el tiempo de duración establecido

Semana o día	Contenidos / Actividad	Modalidad (presencial, virtual sincrónica o virtual asincrónica)	Carga horaria
1	Introducción general	presencial y virtual sincrónica	3
2	Introducción a python, numpy, pytorch.	presencial y virtual sincrónica	7
3	Estadística.	presencial y virtual sincrónica	5
4	Modelos lineales	presencial y virtual sincrónica	5
5	Árboles y bosques aleatorios	presencial y virtual sincrónica	5
6	Clustering y reducción de la dimensionalidad.	presencial y virtual sincrónica	5
7	Máquinas de soporte vectorial.	presencial y virtual sincrónica	5
8	Redes neuronales. Entrenamiento.	presencial y virtual sincrónica	5
9	Redes convolucionales.	presencial y virtual sincrónica	5
10	Redes recurrentes.	presencial y virtual sincrónica	5
11	Aprendizaje profundo. UNET.	presencial y virtual sincrónica	5
12	Transformers. Atención. Modelos de lenguaje	presencial y virtual sincrónica	5
Total horas			60

21. Equipo Docente a cargo:

Señalar nombres y apellidos de los docentes a cargo, títulos de posgrado que poseen y las funciones que cumplirán dentro del equipo: Docente/s responsables, Docente/s – Tutor/es; en los casos en que la complejidad y la extensión de la actividad lo requiera, se podrán prever los roles de Coordinador académico, responsable técnico). Incluir CV resumido de los dictantes (Titulación, actividad docente de grado y posgrado, actividad de investigación, formación de recursos humanos.

Dr. Manuel Pulido (Docente coordinador y dictante).

Dr. Emilio Angelina (Docente dictante)

Dra Magdalena Lucini (Docente dictante)

Dr. Sonia Mariño (Docente dictante)

Dr. Maria Ines Pisarello (Docente dictante)

Colaboradores en el armado y dictado de las actividades prácticas:

Dr. Claudio Rodas

Dr. Sebastian Samoluk

Mgter. Federico Agostini

Lic. Luis Duarte

Lic. Sebastian Filipigh

Prof. Leonardo Gómez Chávez

Lic. Ernesto Rafael Perez

Lic. Santiago Rosa

Ing. Sebastián Suaid

Cabe destacar que más allá del dictado del curso en sí, este curso requiere del armado de un importante material. En particular del armado de las bases de datos, preprocesamiento de los datos, armado de guías de trabajos prácticos en “cuadernos” de jupyter notebooks con códigos que sirvan de ejemplos base del módulo correspondiente.

22. Infraestructura y equipamiento necesario:

Consignar las instalaciones, equipos y apoyos didácticos, recursos materiales y técnicos requeridos para el dictado de la actividad formativa).

Se necesitará acceso a un aula de Informática de FaCENA con computadoras con acceso a internet. También se recomienda a los interesados asistan al curso con una computadora personal portátil (software requerido python, pytorch y demás librerías de machine learning en python de código abierto). Las prácticas serán llevadas a cabo en google colab y en el servidor de machine learning del CECONEA.

Se requiere de un cañón para presentaciones de los teóricos y los prácticos.

23. Fuente/s de financiamiento:

Consignar con qué recursos se financiará la actividad formativa

Se prevé el cobro de un arancel de \$12.000 a docentes y estudiantes de postgrado de la facultad. Para egresados y profesionales el arancel es de \$18.000.

24. Becas

Indicar cantidad de becas que cubran el costo total del programa que puedan ser ofrecidas en el marco de los procesos de internacionalización de la universidad o para mejorar las posibilidades de inserción laboral de recientes graduados.

Se prevé el otorgamiento de hasta tres becas.

CONFORMIDAD

El cuerpo docente presta su conformidad para participar en el dictado del curso que aquí se presenta.