Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Ref. Expte. Nº 797/2021

Ciudad Autónoma de Buenos Aires,31/05/21

VISTO:

La nota presentada por la Dirección del Departamento de Física, mediante la cual eleva la información del curso de posgrado **Laboratorio Avanzado de Datos** para el año 2021,

CONSIDERANDO:

lo actuado por la Comisión de Doctorado,
lo actuado por la Comisión de Posgrado,
lo actuado por este Cuerpo en la sesión realizada en el día de la fecha,
en uso de las atribuciones que le confiere el Artículo 113º del Estatuto Universitario,

EL CONSEJO DIRECTIVO DE LA FACULTAD
DE CIENCIAS EXACTAS Y NATURALES
R E S U E L V E:

ARTÍCULO 1º: Aprobar el nuevo curso de posgrado **Laboratorio Avanzado de Datos** de 96 horas de duración, que será dictado por el Dr. Enzo Tagliazucchi con la colaboración de Sebastián Pinto, Tomás Cicchini, Ariel Berardino.

ARTÍCULO 2°: Aprobar el programa del curso de posgrado **Laboratorio Avanzado de Datos** para su dictado en julio de 2021.

ARTÍCULO 3°: Aprobar un puntaje máximo de cuatro (4) punto para la Carrera del Doctorado.

ARTÍCULO 4°: Disponer que de no mediar modificaciones en el programa, la carga horaria y el arancel, el presente Curso de Posgrado tendrá una vigencia de cinco (5) años a partir de la fecha de la presente Resolución.

ARTÍCULO 5º: Comuníquese a todos los Departamentos Docentes, a la Dirección de Estudiantes y Graduados, a la Biblioteca de la FCEyN y a la Secretaría de Posgrado con copia del programa incluida. Cumplido, archívese.

RESOLUCIÓN CD N.º 0759

Dr. JUAN CARLOS REBOREDA
DECANO

Dr. PABLO J. GROISMAN
Secretario Adjunto de Posgrado
FCEyN - UBA

Formulario para la presentación de Cursos de Posgrado/Doctorado – Res. CD2819/18 - ANEXO 1

Información académica

Año de presentación (*)

2021

1-a-

Departamento docente que inicia el tramite:
Física
Nombre del curso:
Laboratorio Avanzado de Datos
Nombre, Cargo y Título del docente responsable:
Dr. **Enzo Tagliazucchi**, Prof. Adjunto dedicación parcial
En caso de dictarse en paralelo con una materia de grado, nombre de la misma:
Laboratorio de Datos
Nombre y Título de los docentes que colaboran con el dictado del curso (*) (*):
Sebastian Pinto (Ay. 1era, DF), **Tomás Cicchini** (Ay. 1era, DF), **Ariel Berardino** (Ay. 2da, DF)
Fecha propuesta para el primer dictado luego de la aprobación:
1er Cuatrimestre 2021

Duración:

Duración total en horas	96
Duración en semanas	16

Distribución carga horaria:

Número de horas de clases teóricas	3
Número de horas de clases de problemas	
Número de horas de trabajos de laboratorio	3
Número de horas de trabajo de campo	
Número de horas de seminarios	

Forma de evaluación:
Entrega de ejercicios obligatorios y trabajo final
Lugar propuesto para el dictado (departamento, laboratorio, campo, etc.):
Laboratorio de computación

Puntaje propuesto para la carrera de doctorado:	4

Número de alumnos:	Mínimo: 0	Máximo:100

Audiencia a quien está dirigido el curso:
Estudiantes de doctorado en Cs. Físicas u otras carreras de Exactas, con interés en familiarizarse con los métodos y aplicaciones de la Cs. de Datos. Materia optativa de grado para Lic. en Cs.

Físicas Materia obligatoria de grado para la
Lic. en Cs. de Datos

Necesidades materiales del curso:
Computadora para cursada a distancia

1-b-

Programa analítico del curso con Bibliografía (puede adjuntarse en hojas separadas):
Primera parte: análisis exploratorio de datos numéricos Introducción a la ciencia de datos. Consideraciones prácticas y éticas. Repaso Python, Google Colab, herramientas para la materia. Distintos tipos de datos en ciencia de datos. Procesamiento y preparación de datos numéricos. Estandarización, normalización, detección y remoción de outliers, datos faltantes. Estadística descriptiva. Medias, medianas, cuartiles, desvío estándar, error estándar. Distribuciones. Rudimentos de probabilidad. Visualización de datos. Tipos de plots básicos y cuando usarlos: scatter, barras, boxplots, violinplots, tortas, y otros. Segunda parte: modelos de regresión Regresión lineal, cuadrados mínimos. Introducción a scikit-learn. Regresión lineal, cuadrados mínimos. Regresión lineal, polinomios. Concepto de overfiteado. Regresión logística. Clasificación vs. regresión. Regresión lineal y logística como algortimos de ML. Conceptos básicos de ML: función de costo, optimización, gradient descent Tercera parte: modelos de clasificación Test set-train set. Cross-validation. Métricas para la medición de performance en regresión y clasificación. Label shuffling. Clasificador lineal. Clasificación basada en instancias (KNN) Árboles de decisión. Ensembles de árboles aleatorios (random forest) Support vector machines. Kernels.

Cuarta parte: clustering y reducción de la dimensionalidad
Clustering de datos y reducción de la dimensionalidad (ejemplo y análisis: PCA)

Clustering de datos y reducción de la dimensionalidad (ejemplo y análisis: kmeans, kmeans jerárquico)

Quinta parte: obtención de datos, datos no numéricos
Uso de APIs, obtención de keys, requests, distintos

ejemplos Scrapeo de páginas web.

Preprocesado y limpieza de datos no numéricos, principalmente texto.

Preprocesado de datos de texto (segunda parte): matriz de frecuencia, TF-IDF, nube de palabras, análisis de emotividad.

Sexta parte: introducción al procesamiento del lenguaje natural. Clasificación de documentos (KNN, Bayes)

Detección de tópicos (LSA, NMF, LDA)

Ejemplos de adquisición y visualización de otros tipos de datos (datos geoespaciales, visualización)

Bibliografía

Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.

VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. ” O’Reilly Media, Inc.”.

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

Wilke, C. O. (2019). Fundamentals of data visualization: a primer on making informative and compelling figures. O’Reilly Media.

1-c-

Actividades prácticas propuestas (puede adjuntarse en hojas separadas):
Entrega de ejercicios obligatorios relacionados con las primeras cuatro partes de la materia (individual) Entrega de un trabajo práctico final para la aprobación de la materia (trabajo práctico grupal)

(*) Todos los cursos tendrán una validez de 5 años
(*)(*) Las actualizaciones de los docentes colaboradores son informados por la Dirección departamental al inicio de cada dictado del curso

Firma Subcomisión Doctorado	Firma del docente responsable

E-mail y teléfono del docente

tagliazucchi.enzo@df.u
ba.ar 1528516819

responsable

Formulario para la presentación de Cursos de Posgrado/Doctorado - Res. CD2819/18 - ANEXO 2

Solicitud de Financiación

Año de presentación (*)
2021

Departamento docente que inicia el tramite:
Física
Nombre del curso:
Laboratorio Avanzado de Datos
Nombre y Título del docente responsable:

Costo propuesto del curso por alumno (*):

Justificación del monto propuesto:

(*) Las excepciones aplicables para cada alumno serán consistentes con la reglamentación del Consejo Directivo que regula los aranceles y excepciones (Res. CD 484/13). El docente responsable del curso solicitará las excepciones por nota al consejo directico a través de Mesa de Entradas.