

Universidad de Buenos Aires

Facultad de Ciencias Económicas

DEPARTAMENTO DE ECONOMÍA

Carrera: Licenciatura en Economía

Asignatura: Big Data y Machine Learning

Cátedra:

1) ENCUADRE GENERAL

Motivación, Objetivos y Dinámica del curso. Expectativas. Introducción: Definiciones de Big Data, Machine Learning, Ciencia de Datos, Inteligencia Artificial, Estadística y Deep Learning. Machine Learning: Tipos de Aprendizaje: Supervisado, No Supervisado, Aprendizaje por Refuerzo. Como Piensan las Máquinas. Las V del Big Data. Ciencias Computacionales Tradicionales Vs Inteligencia Artificial.

Python: Introducción a la programación en Python. Tipos de Datos, Operaciones y Funciones. Variables. Condicional y For Loops. Módulos: OS, Pandas y Numpy. Pandas: Dataframes, Funciones y Atributos. Análisis Exploratorio de Datos, Tablas Pivot y Visualización de Datos. Web Scrapping. Fusión (merge) de dataframes.

Machine Learning, Aprendizaje Supervisado: Regresión Lineal Simple, Regresión Lineal Múltiple, Función de Pérdida y Costo. Regularización Lasso y Ridge. Clasificación: Regresión Logística, Árboles de Decisión, Ensemble de Árboles.

Machine Learning, Aprendizaje No Supervisado: Cluster, Análisis de Componentes Principales y Detección de Anomalías.

Introducción al Aprendizaje por Refuerzo y al Deep Learning.

Trabajo Práctico.

a) Razones que justifican la inclusión de la asignatura dentro del plan de estudios. Su importancia en la formación profesional

El Big Data y el Machine Learning, son las ciencias del hoy. El mundo actual, en donde la generación de datos y la inteligencia artificial se han vuelto en la materia corriente de todos nuestros días, nos involucra a todos y está en todas partes. En el mundo de hoy, será impensable no entender conceptos básicos de Machine Learning, modelos, aprendizajes, programación y en última instancia como piensan las máquinas que cada vez toman más decisiones en nuestro día a día.

b) Ubicación de la asignatura en el currículum y requisitos para su estudio.

La asignatura se encuentra como asignatura electiva en la curricula del actual plan de estudios.

c) Objetivos del aprendizaje (Misión de la asignatura)

- § Comprender los conceptos actuales relacionados al Machine Learning y el Big Data
- § Aprender nociones básicas de programación en Python
- § Comprender métodos de Aprendizaje Supervisado y poder llevarlos a la práctica.
- § Comprender métodos de Aprendizaje No Supervisado y poder llevarlos a la práctica.
- § Comprender métodos de Aprendizaje por Refuerzo y Deep Learning.

2) PROGRAMA ANALITICO

UNIDAD TEMATICA Nro. 1 - Introducción de conceptos.

Objetivo de aprendizaje:

Comprender y desarrollar los distintos conceptos de la materia, breve reseña histórica sobre como evolucionó el almacenamiento, la producción y explotación de los datos, el Big Data y el Machine Learning, La inteligencia Artificial, los distintos paradigmas y tipos de aprendizajes.

Temas a desarrollar:

- § Introducción
- § Conceptos

UNIDAD TEMATICA Nro. 2 - Generalidades del Machine Learning y el Big Data

Objetivo de aprendizaje:

Desarrollar en profundidad los conceptos enunciados en la unidad anterior.

Temas a desarrollar:

- § Big Data
- § Machine Learning
- § Deep Learning

§ Ciencia de Datos

§ Inteligencia Artificial

UNIDAD TEMATICA Nro. 3 - Introducción a la programación con Python

Objetivo de aprendizaje:

Comenzar a programar con Python y desarrollar la capacidad escribir e interpretar código de programación.

Temas a desarrollar:

§ Tipos de Datos, Variables, Listas, Diccionarios

§ Condicional IF, Bucle FOR, Funciones y Módulos.

§ Introducción a Pandas, Dataframes, Atributos, Funciones y Métodos.

§ Dataframes.

UNIDAD TEMATICA Nro. 4 - Aprendizaje Supervisado: Regresión Lineal y Logística

Objetivo de aprendizaje:

Entender el Aprendizaje Supervisado. Llegar a comprender un ejemplo de regresión lineal simple y múltiple, y una regresión logística para la clasificación binaria.

Temas a desarrollar:

§ Regresión Lineal Simple.

§ Regresión Lineal Múltiple

§ Regresión Logística

UNIDAD TEMATICA Nro. 5 - Aprendizaje Supervisado: Árboles de Decisión

Objetivo de aprendizaje:

Comprender como con Machine Learning Supervisado se puede hacer una clasificación múltiple

Temas a desarrollar:

§ Árboles de Decisión, Nodos, Ramas, Subárboles.

§ Pérdida de Ganancia y Entropía.

§ Coeficiente de Gini y Poda.

UNIDAD TEMATICA Nro. 6 - Aprendizaje No Supervisado: Cluster

Objetivo de aprendizaje:

Comprender como funciona un algoritmo de agrupamiento. K-means y Cluster Jerárquico.

Temas a desarrollar:

§ K-means - Conceptos de Centroides

§ Dendrograma y Cluster Jerárquico.

UNIDAD TEMATICA Nro. 7 - Aprendizaje No Supervisado: Análisis de Componentes Principales y Detección de Anomalías

Objetivo de aprendizaje:

Entender el concepto de Análisis de Componentes Principales y Detección de Anomalías.

Temas a desarrollar:

§ Análisis de Componentes Principales.

§ Bosque de Aislamiento (Isolation Forest)

§ (Factor de Anomalía Local (Local Outlier Factor)

§ (Vector de Soporte por Máquina de Una Clase (One Class SVM)

3) BIBLIOGRAFIA

a) BIBLIOGRAFÍA OBLIGATORIA

La Era de las Máquinas (Varios Autores, Creative Commons Licence, 2020)

Inteligencia Artificial (Gallego, Soto, Aguilera, Sprock et al, 2014)

Libro Vivo de Ciencias de Datos (Cases, 2019)

Análisis Exploratorio de Datos (Figueras, Gargallo, 2003)

b) BIBLIOGRAFIA AMPLIATORIA

Análisis Predictivo: Técnicas y Modelos... (Timón, Fontes, 2017)

Series Temporales (Catalán, S/F)

Introducción a la Limpieza y Análisis de Datos (Subirats, Perez, Calvo, 2019)

Aprendizaje Automático para el Flujo de Datos (Ramos Fernandez, 2019)

Python Data Science Handbook (VanderPlas, 2022)

4) METODOS DE CONDUCCIÓN DEL APRENDIZAJE

a) Objetivos generales a cumplir en los cursos de promoción

El objetivo general del curso es llegar a entender los conceptos básicos del Machine Learning y el Big Data. Lograr programar en Python para poder hacer un Análisis Exploratorio de Datos y ejecutar una secuencia de códigos de un script de aprendizaje supervisado y no supervisado. Llegar a hacer una prueba de concepto de un algoritmo de Machine Learning.

b) Metodología del proceso enseñanza - aprendizaje

La metodología es centrarse en un proceso teórico-práctico de estudio y ejecución de código, y la comprensión de como funcionan los distintos tipos de aprendizajes y algoritmos para poder aplicarlos en distintas pruebas de conceptos.

c) Dinámica del dictado de las clases

Las clases serán presenciales y virtuales alternadas, por Teams y La Facultad de Economía de la UBA.

5) METODOS DE EVALUACION

a) Cursos presenciales y semipresenciales (cursos virtuales y a distancia)

La materia será de cursada mixta (virtual y presencial).

b) Régimen de exámenes finales, intensivos, magistrales y libres

Los alumnos serán evaluados en grupo de hasta 5 personas, con un trabajo práctico final en donde se desarrollara una prueba de concepto desde la lectura de un dataset, el Análisis Exploratorio de Datos, y la aplicación de uno de los algoritmos vistos en la cursada. El trabajo se aprobará con una nota mínima de 4 (cuatro) puntos, que en el caso de no alcanzarse, se solicitará al grupo agregar mayor contenido. La instancia de recuperatorio también podrá ser utilizada para aquellos casos que tengan calificaciones iguales o superiores a 4 (cuatro) y menores a 7 (siete). Esta nota será el concepto de parcial y final de la materia, con lo cual la materia será promocional.

La calificación obtenida en el examen recuperatorio reemplazará a la nota del parcial que se recupera.

c) Criterio de confección del promedio de notas finales

En los casos en que fuere necesario expresar en número entero el promedio de notas parciales o de estas y el examen parcial, se aplicará el número entero superior si la fracción fuere de 0.50 puntos o más y el número entero inferior si fuere de 0.49 o menos. Cuando la nota fuese de 3.01 a 3.99 se calificará con 3 (tres) puntos. (Resolución CS 4994/93)

