

## PROGRAMA DE ASIGNATURA

### ACTIVIDAD CURRICULAR:

#### Ciencia de Datos

**Código:** 95-2577

**Área:** Científico - Técnica

**Bloque:** Tecnología Aplicada

**Nivel:** 5 º

**Tipo:** Electiva

**Modalidad:** Cuatrimestral

**Carga Horaria Total:** 60hs.reloj //80hs. cátedra

**Carga Horaria Semanal:** 4hs. reloj// 5hs. cátedra

### COMPOSICIÓN DEL EQUIPO DOCENTE:

**Docente:** Ing. Ms . Martin Palazzo

**Ayudante:** Ing. Ms. Agustín Velázquez

### FUNDAMENTACIÓN DE LA ASIGNATURA:

El objetivo de esta materia es ofrecer a los estudiantes una doble competencia, tanto de método científico como profesional, con herramientas para la utilización en industrias modernas, dinámicas y de alta complejidad. La industria en general se encuentra cambiando drásticamente desde el nacimiento de internet. La explosión de datos abre nuevas oportunidades a los ingenieros industriales para desarrollarse

como profesionales del análisis de datos, optimización de sistemas y toma de decisiones en entornos de alta complejidad. Se busca formar al futuro ingeniero con las siguientes competencias:

Competencia metodológica en modelización: ofreciendo formación en enfoques de aprendizaje estadístico, ciencia de datos, reconocimiento de patrones e inteligencia artificial en contextos de Ingeniería Industrial (conocida como la cuarta revolución industrial según el World Economic Forum).

Las competencias desarrolladas por la formación se buscan en sectores industriales muy diversos (como lo demuestra la integración profesional de los estudiantes): energía, automotriz, química y petroquímica, logística, empresas especializadas en servicios e ingeniería.

Prepara a los estudiantes para a sus estudios de posgrado, alineado con la propuesta académica de la maestría “ingeniería en sistemas complejos” que el departamento co-tutela con la UTT y con cualquier oferta académica que incluya análisis de datos e Inteligencia Artificial.

## **OBJETIVOS:**

### **Objetivos Generales:**

- Aprender a analizar datos con herramientas estadísticas (estadística descriptiva).
- Comprender los principales algoritmos de Aprendizaje Automático (Machine Learning) y su utilidad en cada caso.
- Iniciar las primeras prácticas con herramientas de programación avanzada (Python).
- Detectar oportunidades de aplicación en la industria en el contexto de la 4ta revolución industrial.
- Formar futuros ingenieros con ventajas competitivas frente al resto de los egresados de otras universidades.
- Aprender a implementar modelos de aprendizaje automático en entornos de producción.

### **Objetivos específicos:**

- Comprender las bases de la Ciencia de Datos y el Aprendizaje Automático y sus aplicaciones en entornos de Ingeniería Industrial.
- Utilizar un lenguaje de programación (Python) para obtener, limpiar y procesar datos.
- Utilizar herramientas de gestión de datos para almacenar y procesar datos localmente y en infraestructuras en la nube.
- Aplicar métodos estadísticos y de visualización para el análisis exploratorio de datos.

- Aplicar modelos de aprendizaje automático para la resolución de problemas prácticos.
- Comunicar efectivamente los resultados de los análisis a través del uso de estadística descriptiva y técnicas de visualización.
- Poder implementar un modelo de aprendizaje automático en datos provenientes un contexto de Ingeniería Industrial.

## CONTENIDOS:

### a) Contenidos Mínimos:

- Análisis Exploratorio de Datos
- Procesamiento y manejo de datos
- Visualización
- Modelos de clasificación
- Modelos de regresión
- Reducción de la dimensionalidad
- Redes neuronales profundas

### b) Contenidos Analíticos:

#### **Unidad 1: Introducción a Ciencia de Datos y Construcción de modelos**

Conceptos básicos: ciencia de datos, aprendizaje automático, reconocimiento de patrones, aprendizaje estadístico. Teoría del aprendizaje. Aprendizaje Supervisado, semi supervisado, no supervisado, aprendizaje por refuerzos. Función objetivo, instancias, clases, atributos. Hipótesis, espacio de hipótesis. Error fuera de la muestra. Sobreajuste. Validación del modelo.

#### **Unidad 2: Análisis exploratorio de datos**

Base de datos (consultas). Servicios en la nube (procesamiento y almacenamiento). Pre-procesamiento. Transformaciones de datos (normalización, auto-escalado). Generación y selección de características. Visualización de datos. Clustering. Valores atípicos. Separación de datos (entrenamiento, validación y prueba).

#### **Unidad 3: Modelos Lineales**

Regresión lineal (simple y múltiple). Regresión logística. Optimización. Método del gradiente descendente. Regularización (L1, L2, Red elástica). Clasificadores lineales. Representación de hipótesis. Regla de decisión. Detección de anomalías.

**Unidad 4: Reducción de Dimensionalidad**

Técnicas de visualización de datos de alta dimensionalidad. Análisis y selección de componentes principales. Reconstrucción de representaciones comprimidas. Eliminación Recursiva de características (RFE).

**Unidad 5: Introducción a las Redes Neuronales**

Hipótesis no lineales. Regresión. Clasificación multiclase. Funciones de costo. Propagación hacia atrás de errores (backpropagation). Técnicas de inicialización. Regularización.

**Unidad 6: Aprendizaje Profundo Deep Learning**

Redes Neuronales Profundas. Redes Neuronales Convolucionales (Convolutional Neural Networks). Encajes (Embeddings). Redes Neuronales Recurrentes. Regularización en redes neuronales profundas.

**DISTRIBUCIÓN DE LA CARGA HORARIA ENTRE ACTIVIDADES TEÓRICAS Y PRÁCTICAS:**

Tipo de Actividad	Carga Horaria Total en Hs. Reloj	Carga Horaria Total en Hs. Cátedra
Teórica	30	38
Formación Práctica(Total)	20	25
Formación Experimental	4	5
Resolución de Problemas	4	5
Proyectos y Diseño	2	3
Práctica Supervisada	0	0

**ESTRATEGIAS METODOLÓGICAS:**

a) Modalidades de Enseñanza empleadas según tipo de actividad (Teórica-Práctica):

El curso se desarrolla mediante el dictado de clases teóricas y clases prácticas. Se realiza una presentación de introducción teórica al tema que luego se asocia a un proyecto práctico en cual el alumno empieza a trabajar. Por cada tema teórico habrá una implementación práctica y una serie de ejercicios que relacionen los tópicos de la materia con Ingeniería Industrial. Con el avance del proyecto se abordan las restantes temáticas y dificultades. Las clases prácticas tienen un espacio de trabajo en software junto con trabajo en equipos, los alumnos avanzan con el proyecto realizan consultas y reciben consejos y ampliación del tema por parte de los profesores.

**b) Recursos didácticos para el desarrollo de las distintas actividades:**

- Apuntes y presentaciones como guía de trabajo y estudio para el alumno.
- Guía de Ejercicios
- Medios audiovisuales de exposición en clase.
- Entorno de desarrollo integrado (IDE) con aplicación en Python (Jupyter, Spyder, IPython, etc).
- Bases de datos públicas.

**EVALUACIÓN:**

**a) Modalidad:**

Cantidad de instancias de evaluación:1(una)

Cantidad de recuperatorios:2(dos)

Consistirá en un trabajo donde se tenga que desarrollar el planteo de hipótesis, desarrollo de un análisis de datos e implementación de modelos de Aprendizaje Automático utilizando bases de datos públicas. La instancia de evaluación será basada en los temas vistos en las clases hasta el momento. La nota final será la nota del trabajo integral.

El alumno que obtenga al menos 8 (ocho) en la instancia de evaluación se considerará promocionado, y no deberá pasar por la instancia de final. En este caso aprueba la materia de forma directa. Si no alcanzara el puntaje necesario para promocionar, pero aprueba ambas con 6 ó 7, firma la asignatura y deberá rendir examen final. Según resolución CD Nº 276/17 – 01/03/2017 – Criterio 1A.

**b) Requisitos de regularidad:**

Para la firma de los Trabajos Prácticos, los alumnos deberán haber aprobado la instancia de evaluación: haber presentado en tiempo y forma el trabajo a desarrollar en las fechas que se anuncie el primer día de clase, o bien que hayan sido modificadas al momento de solicitar cada trabajo, o por autorización expresa

de la cátedra en casos particulares. Para poder rendir el Examen Final, el alumno deberá tener firmada su libreta de Trabajos Prácticos

**c) Requisitos de aprobación:**

El Examen Final será escrito, y podrá abarcar todos los temas del Programa Analítico. La nota mínima de aprobación será de 6 (seis).

**Promoción en primera instancia**

- Asistencia mínima del 75%.
- Aprobados ambos parciales con una calificación mínima de 8.

**Promoción simple**

El alumno que haya obtenido al menos 8 la evaluación y no en la otra, pasará a una segunda instancia para poder promocionar. El alumno que entra en Promoción Simple deberá rendir un examen complementario, (si el otro parcial resultó aprobado con 6 ó 7) o un recuperatorio en el caso de no-aprobado. En ambos casos solo tendrá una oportunidad y será dentro del ciclo lectivo. En el primer caso deberá aprobar el complemento y para el segundo, la calificación deberá ser al menos de 8. En ambas situaciones promocionará con nota final igual 8. Si no aprobara el examen complementario, firma la asignatura y deberá rendir examen final.

**ARTICULACIÓN HORIZONTAL Y VERTICAL CON OTRAS ASIGNATURAS:**

Por las herramientas y conceptos básicos que se requieren para esta materia electiva se da una articulación vertical con Probabilidad y Estadística, Análisis Matemático, Análisis Numérico y Calculo Avanzado, Investigación Operativa e Informática I y II (por sus herramientas otorgadas).

Se logra una articulación horizontal con todas aquellas materias en cual sus problemas prácticos en la actividad empresarial requieran el análisis de grandes cantidades de datos. Los casos son diversos, pero pueden surgir en Ciencias de los Materiales, Economía, Comercialización, Procesos Industriales, Finanzas, Logística, Simulación, Sistemas de Gestión, Robótica Industrial, etc.

**CRONOGRAMA ESTIMADO DE CLASES:**

Unidad Temática	Duración en Hs.Cátedra
1	10
2	10
3	20
4	10
5	10
6	10
Parciales	10

**BIBLIOGRAFÍA OBLIGATORIA:**

- C. M. Bishop, Pattern recognition and machine learning. New York: Springer, 2006.
- T. Hastie, R. Tibshirani, J. H. Friedman, The Elements of Statistical Learning, Springer, 2001.
- I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press. 2016.
- T. M. Mitchell, Machine Learning. New York: McGraw-Hill, 1997.

**BIBLIOGRAFÍA COMPLEMENTARIA:**

- Hands on Machine Learning and scikit learn
- S. Theodoridis and K. Koutroumbas, Pattern recognition, 4. ed. Amsterdam: Elsevier Acad. Press, 2009.
- K. P. Murphy, Machine learning: a probabilistic perspective. Cambridge, Mass.: MIT Press, 2012.
- Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, Learning from data: a short course. S.I.: AMLbook.com, 2012.
- P. R. Norvig and S. A. Intelligence, A modern approach. Prentice Hall, 2002.
- D. Koller and N. Friedman, Probabilistic graphical models: principles and techniques. Cambridge, MA: MIT Press, 2009.
- J. Pearl and others, Models, reasoning and inference. Cambridge University Press Cambridge, UK:, 2000.

- G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, vol. 103. New York, NY: Springer New York, 2013.
- V. N. Vapnik, Statistical learning theory. New York: Wiley, 1998.
- G. Casella and R. L. Berger, Statistical inference, 2nd ed. Australia
- E. L. Lehmann and J. P. Romano, Testing statistical hypotheses, 3rd ed. New York: Springer, 2005.
- T. H. Cormen and T. H. Cormen, Eds., Introduction to algorithms, 2nd ed. Cambridge, Mass: MIT Press,

**CORRELATIVAS:**

**Para cursar:**           **Cursadas:** Investigación Operativa.

**Aprobadas:** Análisis Numérico y Cálculo Avanzado, Informática II.

**Para Aprobar:**       **Aprobadas:** Investigación Operativa.