



ASIGNATURA:	DESARROLLO DE MODELOS ANALÍTICOS (MACHINE LEARNING)	CÓDIGO:	
DEPARTAMENTO:	INGENIERÍA EN SISTEMAS DE INFORMACIÓN	CLASE:	Cuatrimestral
ÁREA:	INGENIERÍA EN SISTEMAS DE INFORMACIÓN	HORAS SEM.:	6 hs.
BLOQUE:	ELECTIVAS	HORAS / AÑO:	Reloj 72hs./ Catedra 96hs

Fundamentación:

En los últimos años la cantidad de información que se genera diariamente es similar a la generada en los cien años anteriores. Este enorme volumen de datos no se transforma automáticamente en información que brinde a las empresas una ventaja competitiva en el mercado.

Con el aumento de dispositivos que generan datos constantemente, el Big Data, y hasta las ciudades inteligentes, es necesario contar con herramientas que permitan explotar y convertir esos datos generados en información.

Gracias a las numerosas y poderosas librerías "Open Source" cualquier persona puede crear un modelo analítico en su computadora de escritorio, utilizando un gran volumen de datos y crear un instrumento que le permita a su empresa encontrar patrones de comportamiento de los clientes a través de sus datos y así accionar de manera diferenciada y enfocada.

Estas librerías y los algoritmos utilizados para generar los modelos cambian a una velocidad sin precedentes, pero los pasos necesarios en la creación de un modelo óptimo es el mismo hace décadas. Estos pasos son: recolectar los datos, generar el target, limpiar los datos, dividir en entrenamiento y testeo, entrenar el modelo, analizar su desempeño, utilizar el modelo para accionar. Cada uno de estos pasos tiene sus particularidades y para un científico de datos es importante conocerlos.

Frecuentemente las empresas realizan campañas de marketing, o aplican descuentos, a todos sus clientes por igual; desperdiçando recursos e invadiendo a los clientes con propagandas en las cuales no están interesados. Los modelos analíticos permiten realizar campañas enfocadas, lo que se traduce en tener mejores resultados porque se contacta a los clientes con mayores probabilidades de adquirir el producto.

Objetivos:

Esta materia electiva ha sido diseñada precisamente para que los futuros Ingenieros en Sistemas de Información puedan:



*Universidad Tecnológica Nacional
Facultad Regional Buenos Aires*

Reconocer cuando un modelo analítico es útil para resolver una necesidad del negocio y decidir los posibles cursos de acción basándose en datos.

Disponer de las herramientas y el entorno necesario para poder realizar modelos analíticos sin necesidad de contar con una arquitectura a gran escala.

Contar con los conocimientos generales para poder transitar el proceso de creación de modelos.

Comprender como se debe definir el universo, la muestra y el objetivo de un modelo para obtener la mayor ganancia sin tener sesgos.

Transformar las variables de entrada a los modelos para potenciar el poder de predicción de los mismos.

Decidir cuál es el tipo correcto de algoritmo en cada situación para poder predecir o entender una necesidad del negocio.

Conocer las métricas que se utilizan para poder evaluar la performance de un modelo analítico para saber si el mismo funciona bien y si mantiene el poder predictivo en el tiempo.

Disponer de ejemplos para poder presentar los resultados obtenidos en el modelado a los otros equipos a fin de facilitar el entendimiento del modelo y sus variables significativas.

Saber cómo es la relación entre el área de desarrollo de modelos analíticos y el equipo que realiza las campañas comerciales para poder aumentar los resultados de las campañas realizadas por la empresa.

Distinguir los conceptos de paralelismo, concurrencia, asincronismo, y evento y operación (no) bloqueante para el diseño e implementación de arquitecturas de software.

Reconocer los modelos de concurrencia no tradicionales, como el paradigma de actores, memoria transaccional, modelo orientado a eventos, Promises, Corrutinas, Guilds y estructuras libres de conflicto como CRDTs.

Identificar el impacto en la escalabilidad y mantenibilidad de cada uno de los conceptos abordados.

Desarrollar la capacidad de toma de decisiones de arquitectura relacionadas con las tecnologías y estilos presentados en la materia.

Programa analítico:

Unidad Temática 1 – Introducción a los Modelos Analíticos

Usos prácticos de modelos (Marketing, Banca, Salud, RRHH, etc.). Metodología KDD (Knowledge Discovery in Databases). Roles necesarios en el desarrollo de modelos. Herramientas necesarias para el desarrollo de modelos. Conceptos generales de Python.

Unidad Temática 2 – Establecer Objetivo

Modelos Supervisados y no supervisados. Establecer el universo, filtro, restricciones y ventana de tiempo para el modelado. Estacionalidad. Crear muestra, balancear el universo y el objetivo. Analizar preliminarmente posibles usos del modelo.

Unidad Temática 3 – Transformación de Variables

Encontrar fuentes de información necesaria. Crear variables derivadas: mínimo, promedio, máximo, y variación en distintas ventanas de tiempo. Análisis univariado, valores atípicos, nulos, mínimos, máximos, deciles. Análisis bivariado. Tratamiento de variables categóricas y ordinales. Crear la ABT (Analytical Base Table)

Unidad Temática 4 – Algoritmos disponibles

Conceptos generales sobre regresiones, árboles, Random Forest, Lightgbm and K-Means.



Unidad Temática 5 – Reducción de dimensionalidad

Correlación de Variables. Valores únicos. Utilizar los algoritmos vistos en la unidad anterior para reducir la dimensionalidad. Componentes Principales. Análisis bivariado de las variables importantes. Analizar la significancia de cada variable.

Unidad Temática 6 – Modelado

Dividir el universo en Entrenamiento/Testeo/Validación. Estandarizar variables. Cross Validation. Encontrar los Hiperparámetros óptimos, Grid Search y Random Search.

Unidad Temática 7 – Desempeño del modelo

Análisis del modelo mediante distintas medidas de desempeño: AUC (área bajo la curva), precisión, recall, matriz de confusión, LIFT, KS (Kolmogorov-Smirnov), análisis por deciles, estabilidad en el tiempo. Overfitting/Underfitting. Necesidad de recalibrar un modelo.

Unidad Temática 8 – Presentación de Resultados

Variables tradicionales vs variables importantes de acuerdo al modelo. Variables por Deciles. Posibles cursos de acción. Uso del modelo por el equipo de campañas.

Distribución de carga horaria entre actividades teóricas y prácticas:

Tipo de actividad	Carga horaria total en hs. reloj	Carga horaria total en hs. cátedra
Teórica	16	20
Formación Práctica	56	76
Formación experimental	0	0
Resolución de problemas	0	0
Proyectos de diseño	0	0
Práctica supervisada	0	0
Total	72	96

Articulación Horizontal y vertical con otras materias

La asignatura Desarrollo de Modelos Analíticos se articula en forma vertical con dos asignaturas que la preceden en el plan de estudio, específicamente Paradigmas de Programación y Análisis de Sistemas. Cada estudiante deberá tener cursada estas asignaturas al momento de comenzar la cursada. Ambas materias le darán al alumno conocimientos que serán aplicados en el desarrollo de modelos. El Ciclo de Vida visto en Análisis de Sistemas puede ser comparado con la metodología KDD y muchas de las temáticas de dicha materia también serán útiles, como ser las Técnicas para obtener y documentar información recolectada.

Sin dudas Paradigmas de Programación le permitirá al alumno comprender como funciona Python, la herramienta que se utilizará en la materia, ya que la misma sigue el Paradigma Orientado a Objetos.



La asignatura Desarrollo de Modelos Analíticos se articula en forma horizontal con dos asignaturas del plan de estudio, específicamente Inteligencia de Negocios e Inteligencia Artificial.

Muchas veces los modelos analíticos utilizan como fuente de información un Data Warehouse o Datamarts, comprender como funciona, sus ventajas y sus desventajas le permitirá al alumno potenciar la información disponible en las empresas. El seguimiento del desempeño de un modelo en el tiempo puede ser realizado con un Tablero. Además, algunos hallazgos obtenidos en los modelos pueden ser plasmados en Tableros. Muchas veces los tableros son fuente de información para el analista ya que muestra los indicadores importantes para el negocio, siempre es bueno traducir estos indicadores a variables en el modelado.

Inteligencia Artificial les brinda a los alumnos el conocimiento necesario para diferenciar dato de información, y le da una mirada conceptual de cómo funcionan los modelos de machine learning.

El equipo docente participa de reuniones inter-cátedras convocadas por el Departamento, a fin de generar acuerdos temáticos y de metodologías que faciliten la articulación horizontal y vertical entre las distintas asignaturas.

Cronograma estimado de clases:

Unidad temática	Duración en horas cátedra
1	12
2	12
3	12
4	12
5	12
6	12
7	12
8	12

Bibliografía:

- Sebastian Raschka (2015). Python Machine Learning .
- Ian H. Witten y Eibe Frank (2005) Data Mining Practical Machine Learning Tools and Techniques.
- Berry, Michael J. A. (2004) Data mining techniques: for marketing, sales, and customer.
- Jake VanderPlas (2016). Python Data Science Handbook: Essential Tools for Working with Data 1st Edición.
- Sunila Gollapudi (2016) Practical Machine Learning.
- Aurélien Géron (2019) Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition.
- Tanay Agrawal (2020) Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient.
- Susan E. McGregor (2021) Practical Python Data Wrangling and Data Quality.
- Sebastian Raschka, Vahid Mirjalili (2019) Python Machine Learning - Third Edition.
- David Paper (2019) Hands-on Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python.



*Universidad Tecnológica Nacional
Facultad Regional Buenos Aires*

PÁGINAS WEB DE INTERÉS

- Python <https://www.python.org/>
- Scikit Learn. <https://scikit-learn.org/stable/>
- Anaconda <https://www.anaconda.com/products/individual>
- Stack overflow. <https://stackoverflow.com/>
- GitHub. <https://github.com/>
- Towards data Science. <https://towardsdatascience.com/>
- LightGBM <https://lightgbm.readthedocs.io/en/latest/>
- Apache Spark <https://spark.apache.org/>
- Databricks <https://databricks.com/documentation>

Correlativas:

PARA CURSAR:

Cursadas: Redes de Información
Administración de Recursos
Simulación
Ingeniería de Software

Aprobadas: Diseño de Sistemas
Sistemas Operativos
Gestión de Datos

PARA RENDIR:

Aprobadas: Redes de Información
Administración de Recursos
Simulación
Ingeniera en Software