



Universidad Tecnológica Nacional
Facultad Regional Buenos Aires

PROGRAMA ANALÍTICO DE ASIGNATURA

DEPARTAMENTO: Ingeniería en Sistemas de Información

CARRERA: Ingeniería en Sistemas de Información

NOMBRE DE LA ACTIVIDAD CURRICULAR: Desarrollo de Modelos Analíticos

Año Académico: 2023

Área: Sistemas Inteligentes

Bloque: Electivas

Nivel: 5º

Tipo: Electiva

Modalidad: Cuatrimestral

Cargas horarias totales:

<i>Horas reloj</i>	<i>Horas cátedra</i>	<i>Horas cátedra semanales</i>
72	96	6

COMPOSICIÓN DEL EQUIPO DE CÁTEDRA:

Profesor Adjunto: Ing. Sebastián Barbone

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.



*Universidad Tecnológica Nacional
Facultad Regional Buenos Aires*

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; desperdiciando 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:

- 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 cómo 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.



*Universidad Tecnológica Nacional
Facultad Regional Buenos Aires*

CONTENIDOS

Contenidos analíticos

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 & 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 re calibrar 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.



*Universidad Tecnológica Nacional
Facultad Regional Buenos Aires*

BIBLIOGRAFÍA OBLIGATORIA

- Agrawa, T.I (2020) Hyperparameter Optimization in Machine Learning: Make Your Machine Learning and Deep Learning Models More Efficient. Ed. Apress.
- Berry, Michael J. A. (2004) Data mining techniques: for marketing, sales, and customer. Ed. Wiley.
- Gollapud, S. (2016) Practical Machine Learning. Ed. Packt.
- Géron, A. (2019) Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition. Ed. O'Reilly.
- McGregor, S. (2021) Practical Python Data Wrangling and Data Quality. Ed. O'Reilly.
- Pape, D. (2019) Hands-on Scikit-Learn for Machine Learning Applications: Data Science Fundamentals with Python. Ed. Apress.
- Raschka S. Vahid Mirjalili (2019) Python Machine Learning - Third Edition. Ed. Packt.
- Vander Plas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data 1st Edición. Ed. O'Reilly.
- Witten, I. y Eibe Frank (2005) Data Mining Practical Machine Learning Tools and Techniques. Ed. Elsevier.

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 y rendir

- Cursadas:
 - Redes de Datos
 - Administración de Sistemas de Información
 - Simulación



Universidad Tecnológica Nacional
Facultad Regional Buenos Aires

- Ingeniería y Calidad de Software
- Aprobadas:
 - Diseño de Sistemas de Información
 - Sistemas Operativos
 - Bases de Datos