

Tecnológico de Monterrey Santa Fe

Del mundo físico al mundo digital con ciencia de datos (Semana i 2019)
Profesor: Saul Juárez Ordoñez



Objetivos

El objetivo principal de esta semana i es ayudarte a desarrollar pensamiento matemático, estadístico y computacional al abordar un problema de fraude crediticio con herramientas de aprendizaje automático (machine learning). Al finalizar esta semana, tendrás conocimientos generales de algoritmos de clasificación para manejar bases de datos de tu interés y extraer información relevante para una mejor toma de decisiones.

Habilidades específicas a desarrollar

Habilidades de programación cuantitativas y básicas: exploración, procesamiento, análisis estadístico de datos e implementación de algoritmos de clasificación con el lenguaje de programación Python y la plataforma Jupyter. Habilidades de comunicación efectiva oral y escrita.

Descripción de la actividad

Dada una base de datos de transacciones bancarias anonimizada, realizarás

1. Un profundo análisis exploratorio de datos
2. Entrenarás y probarás algoritmos de clasificación de aprendizaje automático para detectar fraude crediticio.

La guía detallada de la actividad se encuentra en los *jupyter notebooks* del caso de estudio que hemos llamado **Fighting Fraud**. Hemos creado dos versiones del caso. Trabaja la versión de tu elección dependiendo de tu nivel de Python: **Beginner Version** y **Advanced Version**. La descripción correspondiente está en el respectivo *notebook* que puedes encontrar en nuestro [repositorio de GitHub](#).

Temario

1. Herramienta Anaconda Navigator
 - a. Instalación del ambiente de trabajo
 - b. Familiarización con la plataforma Jupyter
2. Introducción a Python con Jupyter

- a. Conceptos básicos de Python
 - b. Librerías importantes para análisis de datos y aprendizaje automático
3. Caso de estudio: Fraude crediticio
 - a. Exploración de datos (numérica y visual)
 - b. Preprocesamiento (PCA, conjuntos de prueba y de entrenamiento)
 - c. Muestras balanceadas
 - d. Outliers extremos
 - e. t-SNE
4. Comparación de algoritmos de clasificación de aprendizaje automático

Recursos de apoyo

- Tutorial de instalación de Anaconda
- Curso introductorio de Python
- Ejemplos de Jupyter Notebooks con algunos casos de estudio

Contenido adicional

Referencias

- McKinney, W. (2012). Python for data analysis: Data wrangling with Pandas, NumPy, and IPython. " O'Reilly Media, Inc."
- Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.
- Matthes, E. (2015). Python crash course: a hands-on, project-based introduction to programming. No Starch Press.
- Jupyter Documentation: <https://buildmedia.readthedocs.org/media/pdf/jupyter-notebook/latest/jupyter-notebook.pdf>

Criterios de evaluación

Actividades	Porcentaje
Jupyter Notebook del caso resuelto	60%
Presentación oral final	35%
Auto y co-evaluación	5%

Presentación final

- El viernes 1 de noviembre se debe de presentar el Jupyter Notebook del caso resuelto y exponer su solución ante un panel de evaluadores.

Contacto

- Saúl Juárez Ordóñez- sauljz@tec.mx
- Agustín Ubierna Burgos- agustin.ubi@tec.mx