PROGRAMA ABIERTO DE COMPLEMENTACIÓN Y AMPLIACIÓN DE LA CURRICULA DE LA MAESTRÍA 2017

La Maestría en Generación y Análisis de Información Estadística, en el marco de su Programa de Actualización Permanente de las Orientaciones de Estadísticas Económicas, Estadísticas Sociodemográficas y Estadísticas de Opinión y Mercado, en coordinación con la Secretaría de Extensión Universitaria y la Dirección de Posgrado, presenta el Ciclo de Seminarios y Cursos extracurriculares en Estadística para este año lectivo.

Coordinador: Jorge Fernández Bussy

Curso: "Fundamentos de la programación estadística y Data Mining en R. Estadística descriptiva, modelos de regresión y árboles de decisión"

Docente: Germán Rosati

Doctor en Ciencias Sociales (UBA), Magíster en Generación y Análisis de Información Estadística (UNTREF) y Licenciado en Sociología (UBA). Actualmente se desempeña como Analista Experto de Datos en la Subsecretaría de Políticas, Estadísticas y Estudios Laborales del MTEySS de la Nación y como consultor estadístico independiente. Ha sido becario doctoral del CONICET e investigador visitante en la Freie Universität de Berlín. Ha dictado cursos de estadística de grado y posgrado en diversas universidades (USAL, UNGS, UNSAM, UNCuyo).

Presentación y objetivo del curso:

Debido a su carácter de software libre y a la creciente comunidad de usuarios el lenguaje R se ha convertido en algo así como la lingua franca dentro del análisis estadístico. El presente seminario se propone realizar una introducción a algunos conceptos fundamentales de la programación estadística en R. A su vez, se hará énfasis en la implementación de análisis estadísticos básicos (descriptivos y regresiones) en R. A su vez, el curso presentará algunos elementos teóricos de la minería de datos/aprendizaje automático (balance sesgo-variancia, overfitting, etc.) y revisará algunos algoritmos para la estimación de árboles (ID4, C4.5, CART y random forest).

El curso se propone que los alumnos:

- se familiaricen con aspectos relevantes de la programación estadística en lenguaje R
- logren implementar e interpretar análisis estadísticos descriptivos y modelos de regresión en lenguaje R
- incorporen algunos conceptos fundamentales del data mining/aprendizaje automático,
- conozcan generalidades de algunos algoritmos para la generación de árboles de decisión (ID4, C4.5,CART y random forest) y su implementación en lenguaje R,
- logren identificar situaciones de aplicación de este tipo de modelos a problemas de investigación básica y aplicada

Destinatarios: Estudiantes avanzados de carreras de grado y posgrado, técnicos, profesionales, investigadores, docentes y no docentes.

Temario de clases

Unidad 1. Elementos de programación estadística en R. Objetos en R (vectores, matrices, data frames y listas). Estructuras de control (loops –for, while, repeat- if, ifelse). Implementación de funciones ad-hoc. Generación de números aleatorios y distribuciones de probabilidad. Importación y exportación de datos (.csv, .txt, .tab, .sav, etc.).

Unidad 2. Análisis estadístico básico en R. Generación de gráficos y visualización de datos. Estadística descriptiva. Implementación y análisis de modelos de regresión lineal y logística. Funciones plot, lm, glm y predict.

Unidad 3. Nociones básicas de data mining/aprendizaje automático. Tipos de problemas en aprendizaje supervisado: clasificación y regresión. Error de entrenamiento (training error), error de prueba (test error). Sobre-ajuste. Balance entre el sesgo y la variancia de un modelo. Métodos de estimación del error: partición del dataset, validación cruzada. Aplicaciones en R.

Unidad 4. Clasificadores basados en árboles: generalidades. Algoritmos ID4, C4.5 y CART. Partición múltiple y binaria, medidas de pureza de nodos. Crecimiento (growing) y podado (prunning) de árboles de decisión. Balance entre costo y complejidad del árbol. Introducción a los modelos de Random Forest. Aplicaciones en R (paquetes tree, rpart y randomForest).

Bibliografía básica de referencia

Breiman, L., Friedman, J., Stone, C. y Olshen, R. (1984), Classification and Regression Trees, New York: Champan & Hall/CRC.

Breiman, Leo (2001), “Statistical modelling. The two cultures”, Statistical Science, Vol. 16, n°3: 199-215.

Hastie, T.; Tibshirani, R. & Friedman, J. (2009), The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Berlin: Springer.

Hernández Orallo, J., Ramírez Quintana, J., Ramírez, C., (2004). Introducción a la minería de datos, España: Pearsons Editorial.

James, G., Witten, D., Hastie, T., Tibshirani, R. (2013), An Introduction to Statistical Learning with Applications in R, Berlin: Springer.

R Core Development Team, (2000), Introducción a R. Notas sobre R: Un entorno de programación para Análisis de Datos y Gráficos, (disponible en https://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf)

Tetor, P. (2011), R Cookbook. Proven recipes for data analysis, statistics and graphics, New York: O´Reilly.

Requisitos para la cursada y aprobación

Conocimientos básicos de estadística descriptiva y cierta familiaridad con el modelo de regresión lineal y logística. Será útil (pero no absolutamente necesario) alguna experiencia en programación estadística (sea en SPSS, Stata o similar)
Para la aprobación del curso se requiere:

1) un mínimo de asistencia del 80% sobre el total de clases y
2) la entrega y aprobación de una monografía final

Organización del curso:

Modalidad: Presencial
Días y Horario: miércoles de 18 a 22 horas
Lugar de cursada: Centro Cultural Borges, Viamonte y San Martín, Pabellón de las Naciones, 3º piso, Ciudad de Buenos Aires.
Fecha de inicio: 01/03/17
Fecha de finalización: 29/03/17
Cantidad de clases: 5
Total de horas: 20 hs

Arancel: El curso tiene un costo total de **$2.150**
Alumnos regulares, docentes y graduados de la Universidad Nacional de Tres de Febrero reciben una reducción arancelaria del 50%

Informes e Inscripción
A partir del 6 de febrero escribimos a: maestriaestadistica@untref.edu.ar

Maestría en Generación y Análisis de Información Estadística
Centro Cultural Borges, Viamonte y San Martín, Pabellón de las Naciones, 3º piso.
Ciudad de Buenos Aires.