

Planificación de la Asignatura: Modelado Estadístico

Fecha: 23/10/2024 13:02

Código: T1633

Carrera: Tecnicatura Universitaria en Procesamiento y Explotación de Datos

Departamento Académico: Matemática

Docente a cargo:

Correo del docente a cargo: sin datos

Régimen de Dictado: Cuatrimestral 1º Cuatrimestre

Carga Horaria Semanal: 5 horas semanales

Carga Horaria Total: 70 horas

Contenidos Mínimos:

Regresión y correlación lineal. El problema de predicción. Colinealidad múltiple. Transformaciones de variables para cumplir con los supuestos del modelo. Modelos de clasificación. Variables ficticias (dummy variables). Interacción. Regresión logística. Comparación y selección de modelos. Métodos no paramétricos.

Competencias Genéricas:

Competencias Específicas:

Argumentación de aportes marcados en la matriz de competencias:

Correlativas Regulares para cursar:

Comprensión Lectora y Producción Escrita,
Espacio Integrador I,
Probabilidad y Estadística;
Algoritmos y Estructura de Datos
Aspectos Legales del Uso de la Información

Correlativas Aprobadas para cursar:

Informática Básica
Algebra y Calculo

Correlativas Aprobadas para promocionar o rendir el examen final:

No posee

Insercion de la Asignatura en el plan de Estudios:

La materia se desarrolla durante el primer cuatrimestre del segundo año de la carrera Tecnicatura Universitaria en Procesamiento y Explotación de Datos de la Facultad de Ingeniería, de la Universidad Nacional de Entre Ríos.

Se propone como un espacio para comenzar a comprender las características de los diferentes modelos estadísticos, sus aplicaciones y para la interpretación de resultados.

Dentro del plan de estudios la misma es correlativa con Informática Básica y Álgebra y Cálculo por lo que se posee una base sólida para que los estudiantes programen los algoritmos correspondientes a los modelos en softwares estadísticos así cómo puedan interpretar las implicancias teóricas de los diferentes modelos en sus aplicaciones.

Esto genera que la asignatura se proponga como un espacio para comprender teóricamente el funcionamiento de los modelos y tomar desiciones en la praxis sobre el mejor modelo a aplicar según las características de los datos.

Objetivo General:

Que los estudiantes puedan conocer y comprender las características de diferentes Modelos estadísticos para que puedan aplicarlos en forma eficiente y fundamentada.

Objetivos Particulares:

Qué los estudiantes logren:

- Comprender las características de los modelos de regresión y de clasificación.
- Aplicar los algoritmos de los diferentes modelos que se desarrollan en la asignatura.
- Interpretar los resultados arrojados por los diferentes modelos en el contexto del problema.
- Discernir entre el modelo que mejor se adapte a las necesidades de las situaciones planteadas.

Programa Analítico:

Unidad temática 01: Introducción

Modelado Estadístico: Qué es un modelo. Qué características posee un modelo estadístico y cuáles son sus propósitos.

Unidad temática 02: Regresión Lineal Simple

Correlación lineal: Coeficiente de Pearson, cálculo e interpretación. Otros coeficientes (Rho de Spearman, Thau de Kendall, Jackknife) características y usos. Regresión. Tipos de regresión. Regresión lineal simple: Condiciones para la linealidad, modelo de regresión lineal simple, bondad de ajuste.

Unidad temática 03: Regresión Lineal Múltiple

Coeficientes de correlación parcial y múltiple. Modelo de regresión lineal múltiple. Restricciones al modelo de regresión lineal múltiple. Transformaciones de las variables.

Unidad temática 04: El problema de la clasificación y los modelos.

Clasificación supervisada. Particionamiento de datos para entrenamiento y test. Definición de los indicadores de los modelos: Tasa de aciertos, tasa de errores, especificidad y sensibilidad. Curva ROC. La Regresión como clasificador. Clasificador Bayesiano y Discriminante Lineal. Introducción al modelo y ejemplos.

Unidad temática 05: Transformación de Variables.

Cuándo y porqué transformar una variable. Variables ficticias (dummy variables). Tipos de Transformaciones: Transformación logarítmica, Transformación de raíz (cuadrada u otras), Transformación arcoseno o angular, Transformación recíproca, Transformación de Box-Cox.

Unidad temática 06: Modelos no Paramétricos

Tablas de contingencia: construcción e interpretación. Pruebas: χ^2 de Pearson, Binomial, Cohen kappa, Test de Fisher, Kolmogórov-Smirnov, Kruskal-Wallis, Coeficiente de correlación de Spearman: Cálculo e interpretación de resultados. Modelos ANOVA y ANCOVA. Regresión logística Multinomial: Objetivos y supuestos del modelo. Modelo logit y probit. Introducción al modelo y ejemplos.

Metodología Didáctica:

Los contenidos serán propuestos en las clases denominadas de “teoría” donde se desarrollarán los principios teóricos que caracterizan a cada uno de los modelos. Se espera describir a los estudiantes el funcionamiento general de los algoritmos de cada modelo a fin de comprendan en qué situaciones es factible aplicarlos y qué representan los resultados de los mismos. Así mismo se abordarán ejemplos que permitirán conocer las funciones y librerías de R con las que se pueden aplicar dichos modelos y, finalmente se revisarán las métricas para interpretarlas y realizar reflexiones sobre los resultados en virtud de las características del data set y/o de los parámetros.

En las clases de práctica se propondrán una serie de actividades de dificultad incremental en la que los estudiantes podrán profundizar sobre los temas desarrollados y sobre las posibilidades de mejorar los resultados de los modelos. Es el objetivo principal que los estudiantes desarrollen autonomía para indagar en los procesos de análisis y de los diferentes modelos estudiados

Formación Práctica:

Listado de Actividades de Formación Práctica:

Intensidad de la formación práctica

Detalle de la carga horaria total prevista para cada una de las siguientes actividades:

Actividades prácticas que aportan a las competencias específicas en el Nivel de dominio 1: 0 horas

Actividades prácticas que aportan a las competencias específicas en el Nivel de dominio 2: 0 horas

Actividades prácticas que aportan a las competencias específicas en el Nivel de dominio 3: 0 horas

Horas totales de actividades de formación práctica: 0 horas

Metodología de Evaluación Durante el cursado:

Durante el cursado se observarán las participaciones de los estudiantes en las actividades asincrónicas para ofrecerles espacios de validación de lo comprendido puesto que en el desarrollo de la materia deberán ir realizando un resumen de cada uno de los modelos estadísticos trabajados, sus características, sus ventajas y desventajas, así como de los datasets y la finalidad de los mismos.

Además deben seleccionar un data set, describir sus características, definir un objetivo y/o hipótesis de análisis y luego para implementar uno o varios de los modelos según el objetivo del análisis que se propongan.

Este resumen teórico y el análisis deberán ser entregados como tareas en el aula Virtual y defendido durante las últimas dos semanas de cursado. La aprobación de estas tres actividades con una nota no inferior al 60% será la condición de promoción para la materia.

La nota final será el promedio entre los porcentajes alcanzados.

Metodología de Evaluación en Exámenes Finales:

Para los alumnos Regulares:

- Deberán realizar una serie de actividades prácticas que den cuenta de poder aplicar los modelos estadísticos trabajados en clase con el software R, realizando una interpretación de las salidas en cada caso y proponiendo posibles mejoras.
- También deberán contestar un par de consignas teóricas que den cuenta de conocer los contenidos de la materia.

Estas consignas deberán resolverse en un tiempo de entre dos y tres horas. Esta actividad debe conseguir una puntuación del 60% para que el estudiante esté aprobado.

Para los alumnos Libres:

Además de aprobar la instancia de evaluación propia de los estudiantes regulares se le solicitará que, como mínimo, una semana antes de rendir presente un trabajo donde deben seleccionar un data set, describir sus características, definir un objetivo y/o hipótesis de análisis y luego para implementar uno o varios de los modelos según el objetivo del análisis que se propongan. Remitirlo a los docentes evaluadores y aprobarlo con una nota igual o superior al 60%

Condiciones de Regularidad :

Para que un estudiante alcance la regularidad de la materia será necesario que realice todas las actividades asincrónicas de la materia y que presente en las dos últimas semanas de la materia el resumen teórico y La actividad de análisis del data set propuestos para la promoción. Si la nota de al menos uno de estos es inferior al 60 % el estudiante no podrá defender su trabajo para la promoción y, por dicho motivo, deberá recibir las correcciones pertinentes y presentarse en la mesas de finales correspondientes.

Cronograma de parciales durante el primer Cuatrimestre:

Primer Examen Parcial: 26 de Abril de 2024

Segundo Examen Parcial: 04 de Junio de 2024

Recuperatorio 01: 14 de Junio de 2024

Cronograma de parciales durante el segundo Cuatrimestre:

Bibliografía Principal:

Fernando Tellez, C. - Alfonso Morales, M. (2016) Modelos Estadísticos Lineales, con Aplicaciones en R. Ediciones De La U. Colombia 2016 ISBN 9789587624762

Garibaldi, Lucas A. y otros (2019) Modelos estadísticos en lenguaje R Universidad Nacional de Río Negro ISBN 978-987-4960-05-4

Introducción a R:

https://www.datacamp.com/courses/introduccion-a-r/?tap_a=5644-dce66f&tap_s=10907-287229

Parra, F. Vicente, J.A. (2019): Apuntes modulo3: Análisis de Datos Multivariantes I. Master de Data Science y Big Data aplicados a la Economía y a la Administración y Dirección de Empresas. UNED.

Parra, Francisco (2019) Estadística y Machine Learning con R. URL:

<https://bookdown.org/content/2274/portada.html>

Bibliografía Complementaria:

Berridge, D. M. y Crouchley R. (2011). Multivariate Generalized Linear Mixed Models Using R. Boca Raton, EUA

Cayuela, L (2009). Modelos lineales generalizados. Granada: Universidad de Granada.

Fox, J. y Weisberg, S. (2010). Nonlinear Regression and Nonlinear Least Squares in R. An Appendix to An R Companion to Applied Regression.

Gelman, A. y Hill, J. (2007). Data analysis using regression and ultilevel/hierarchical models. Cambridge, UK: Cambridge University Press.

Kleiber, C. y Zeileis, A. (2008). Applied Econometrics with R. Springer-Verlag: New York.

Equipo de Cátedra:

Actividades de Investigación Gestión y Extensión:

Requisitos de admisión para alumnos oyentes:

Infraestructura, equipamiento y recursos necesarios:

Otros: