**ANÁLISIS AVANZADO DE DATOS**

Guía de asignatura

Última actualización: julio de 2021

1. **Información general**

|  |  |
| --- | --- |
| **Nombre de la asignatura** | Análisis Avanzado de Datos |
| **Código** | 32310014 |
| **Tipo de asignatura** | Electiva |
| **Número de créditos** | 2 |
| **Tipo de crédito** | 1A+1B |
| **Horas de trabajo semanal con acompañamiento directo del profesor** | 24 |
| **Horas semanales de trabajo independiente del estudiante** | 72 |
| **Prerrequisitos** | Análisis Estadístico de Datos |
| **Correquisitos** | Ninguno |
| **Horario** | Sábados 7h-10h |
| **Líder de área** |  |
| **Salón** | Edificio El Tiempo |

1. **Información del profesor y monitor**

|  |  |
| --- | --- |
| **Nombre del profesor** | Andrés Pérez Coronado |
| **Perfil profesional** | https://www.linkedin.com/in/andrés-pérez-coronado-63792618/ |
| **Correo electrónico institucional** | Andresg.perez@urosario.edu.co |
| **Lugar y horario de atención** |  |
| **Página web u otros medios** |  |

1. **Resumen y propósitos del curso**

Este curso presenta modelos estadísticos avanzados para el análisis de datos. El curso cubre generalizaciones al modelo de regresión lineal que permiten mayor flexibilidad de estimación. El curso también introduce modelos de datos estructurados y variables latentes, así como modelos para el análisis de causalidad.

El curso busca formar al estudiante en métodos estadísticos avanzados para el análisis de datos, considerando sus fundamentos teóricos y sus aplicaciones.

1. **Conceptos fundamentales**

1. Evaluación de modelos y bootstrap
2. Métodos de suavización: splines y métodos de kernel
3. Modelos generalizados lineales y aditivos
4. Variables latentes y datos estructurados
5. Datos dependientes

1. **Resultados de aprendizaje esperados (RAE)**
2. Evaluar la bondad de un modelo estadístico.
3. Emplear modelos lineales generalizados para la estimación de relaciones entre variables.
4. Identificar situaciones donde el uso de métodos de suavización es adecuado.
5. Estimar modelos estadísticos usando variables latentes y datos estructurados.
6. Modelar usando datos dependientes en la solución de problemas aplicados.
7. **Modalidad del curso**

Presencial.

1. **Estrategias de aprendizaje**

El desarrollo del curso se efectuará por medio de la construcción del fundamento teórico a través de la acción facilitadora del docente y la interacción con los estudiantes, para lo cual el estudiante debe desarrollar una lectura previa del contenido a partir de la segunda sesión, de esta manera podrá aportar a la discusión y a la consolidación del fundamento teórico. Una vez conseguido el fundamento teórico, este será reforzado y contextualizado por medio de laboratorios y casos de estudio. También se estimulará la participación del estudiante mediante trabajos de indagación con la debida sustentación ante el profesor y el curso.

El desarrollo de las sesiones se efectuará con el apoyo de la plataforma e-aulas y notebooks *(jupyter*); en el transcurso del curso se plantearán problemas donde el estudiante deberá identificar los conceptos aplicables a la situación planteada, generando espacios para la indagación, la resolución de situaciones, la confrontación de opiniones y la toma de decisiones.

* Evaluaciones
* Estudio de caso
* Talleres y ejercicios

1. **Actividades de evaluación**

Se recomienda la evaluación por estudio de casos, proyectos, laboratorios. Sin embargo, el docente puede usar otras actividades que deben quedar explicitas en la siguiente tabla. Tenga presente que en la siguiente tabla deben evidenciarse la evaluación de todos los Resultados de Aprendizaje.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Actividad** | **Descripción** | **Porcentaje** | **RAES evaluados** | **Fechas** |
| Proyecto final: Decisiones basadas en datos | Simule un equipo de consultores que se presenta ante la gerencia y la junta directiva de una organización. Adapte su mensaje a sus perfiles, preferencias y visión organizacional para que sea lo más personalizado posible. | 60% | Evaluar la bondad de un modelo estadístico.  Emplear modelos lineales generalizados para la estimación de relaciones entre variables.  Identificar situaciones donde el uso de métodos de suavización es adecuado.  Estimar modelos estadísticos usando variables latentes y datos estructurados.  Modelar usando datos dependientes en la solución de problemas aplicados. | Presentacion última sesión |
| Evaluación de participación en clase | La participación en clase se evaluará en función de la profundidad y la calidad del análisis realizado sobre el contenido del curso y los resultados de los datos presentados a lo largo de las sesiones. Se espera que los estudiantes interactúen activamente con los materiales y proporcionen información e interpretaciones significativas en el repositorio de GitHub. | 40% | Evaluar la bondad de un modelo estadístico.  Emplear modelos lineales generalizados para la estimación de relaciones entre variables.  Identificar situaciones donde el uso de métodos de suavización es adecuado.  Estimar modelos estadísticos usando variables latentes y datos estructurados.  Modelar usando datos dependientes en la solución de problemas aplicados. | Al final de cada sesión |

1. **Programación de actividades**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S** | **Tema** | **Descripción de la actividad** | **Trabajo independiente** | **Recursos que apoyan la actividad** |
| 1 | Análisis exploratorio de datos multivariantes | Medias de centralidad, variabilidad y dependencia lineal multivariante. | Lectura previa, tareas y talleres | [2] 3  [6] 1-2-3 |
| 2 | Análisis de correspondencias y correlaciones | Correlaciones lineales, no paramétricas y análisis de correspondencias. | Lectura previa, tareas y talleres | [2] 4-6-7  [6] 7-8 |
| 3 | Variables latentes y datos estructurados | Modelos mixtos  El algoritmo EM  Análisis de componentes independientes (ICA)  Otras técnicas de reducción de la dimensión,  Datos perdidos e imputación. | Lectura previa, tareas y talleres | [1] 9.6-14.7  [2] 5-8-11-12 |
| 4 | Boostrap, Métodos de suavización, Splines, | The bootstrap, Polinomios a trazos y splines  Splines suavizados y splines multidimensionales | Lectura previa, tareas y talleres | [1] 5.1-5.2-5.4-5.6-7.11-8.2  [2] 9-10  [6] 9-10-11 |
| 5 | Métodos de Kernel, Modelos generalizados | Métodos kernel de suavización  Estimación de la densidad mediante kernels  Regresión logística, LDA  Modelos lineales generalizados  Modelos aditivos generalizados | Lectura previa, tareas y talleres | [1] 6.1-6.6; 4.3, 4.4; 9.1  [2] 13-14 |
| 6 | Datos dependientes | Series temporales I: Definición, estacionariedad, modelos autoregresivos, bootstrapping, validación cruzada | Lectura previa, tareas y talleres | [3] 1-2 |
| 7 | Series temporales II: Tendencias, removiendo tendencias, saltos en series temporales, variables latentes, y series temporales multivariantes | Lectura previa, tareas y talleres | [3] 3-4 |
|  |
| 8 | **Sustentación del Proyecto Final**  **Conclusiones del curso** | | | |

1. **Factores de éxito para este curso**

A continuación, se sugieren una serie de acciones que pueden contribuir, de manera significativa, con el logro de metas y consecuentemente propiciar una experiencia exitosa en este curso:

1. Planificar y organizar el tiempo de trabajo individual que le dedicará al curso
2. Organizar el sitio y los materiales de estudios
3. Tener un grupo de estudio, procurar el apoyo de compañeros
4. Cultivar la disciplina y la constancia, trabajar semanalmente, no permitir que se acumulen temas ni trabajos
5. Realizar constantemente una autoevaluación, determinar si las acciones realizadas son productivas o si por el contrario se debe cambiar de estrategias
6. Asistir a las horas de consulta del profesor, participar en clase, no quedarse nunca con la duda
7. Utilizar los espacios destinados para consultas y resolución de dudas, tales como Sala Gauss y Sala Knuth
8. Propiciar espacios para el descanso y la higiene mental, procurar tener buenos hábitos de sueño
9. Tener presente en todo momento valores como la honestidad y la sinceridad, al final no se trata solo de aprobar un examen, se trata de aprender y adquirir conocimientos. El fraude es un autoengaño.
10. **Bibliografía y recursos**

[1] T. Hastie and R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*. Springer.

[2] D. Peña, *Análisis multivariante de datos*, McGraw-Hill Interamericana de España S.L., 2002.

[3] Daniel Peña, Ruey S. Tsay, Statistical Learning for Big Dependent Data, 2021 John Wiley & Sons Inc.

1. **Bibliografía y recursos complementarios**

[4] Wasserman, Larry. All of Nonparametric Statistics, Springer. ISBN 0-387-25145-6. 2207,

[5] H. Trevor and R. Witten and G. Tibshirani, *Introduction to Statistical Learning with R.* Springer*. 2015.*

[6] Moraga, Paula. (2023). Spatial Statistics for Data Science: Theory and Practice with R. Chapman & Hall/CRC Data Science Series. ISBN 9781032633510.

1. **Acuerdos para el desarrollo del curso**

No está permitido comer o usar dispositivos móviles dentro de clase. No se realizará aproximación de notas al final del semestre. Las notas solo serán cambiadas con base en reclamos OPORTUNOS dentro de los límites de tiempo determinados por el Reglamento Académico. Si por motivos de fuerza mayor el estudiante falta a algún parcial o quiz, deberá seguir el procedimiento regular determinado por el Reglamento Académico para presentar supletorios. No habrá acuerdos informales al respecto. No se eximirá a ningún estudiante de ningún examen.

**PROCESOS DISCIPLINARIOS - FRAUDE EN EVALUACIONES**

Teniendo en cuenta el reglamento formativo-preventivo y disciplinario de la Universidad del Rosario, y la certeza de que las acciones fraudulentas van en contra de los procesos de enseñanza y aprendizaje, cualquier acto corrupto vinculado a esta asignatura será notificado a la secretaría académica correspondiente de manera que se inicie el debido proceso disciplinario. Se recomienda a los estudiantes leer dicho reglamento para conocer las razones, procedimientos y consecuencias que este tipo de acciones pueden ocasionar, así como sus derechos y deberes asociados a este tipo de procedimientos.

La asignatura no tiene ningún tipo de Bono.

Debe consultar:

<http://www.urosario.edu.co/La-Universidad/Documentos-Institucionales/ur/Reglamentos/Reglamento-Academico-de-Pregrado/>

1. **Respeto y no discriminación**

Si tiene alguna discapacidad, sea este visible o no, y requiere algún tipo de apoyo para estar en igualdad de condiciones con los(as) demás estudiantes, por favor informar a su profesor(a) para que puedan realizarse ajustes razonables al curso a la mayor brevedad posible. De igual forma, si no cuenta con los recursos tecnológicos requeridos para el desarrollo del curso, por favor informe de manera oportuna a la Secretaría Académica de su programa o a la Dirección de Estudiantes, de manera que se pueda atender a tiempo su requerimiento.

Recuerde que es deber de todas las personas respetar los derechos de quienes hacen parte de la comunidad Rosarista. Cualquier situación de acoso, acoso sexual, discriminación o matoneo, sea presencial o virtual, es inaceptable. Quien se sienta en alguna de estas situaciones puede denunciar su ocurrencia contactando al equipo de la Coordinación de Psicología y Calidad de Vida de la Decanatura del Medio Universitario (Teléfono o WhatsApp 322 2485756).