

Materia:	Analítica de Datos
Requisitos:	(SIS) Inferencia Estadística y Curso Integrador Uno
Programa – Semestre:	
Período académico:	2018 - 2
Intensidad semanal:	3 horas
Créditos:	3

Descripción

El uso inteligente de los datos se ha convertido en uno de los motores de empoderamiento de las industrias. Les ha permitido conocer mejor a sus clientes, mejorar su productividad, generar innovación, y alcanzar nuevos niveles de competitividad.

En segmentos como el financiero, telecomunicaciones o retail, entre muchos otros, la analítica de datos facilita a las empresas la comprensión de su situación actual y de las oportunidades de mejora, a partir de la adquisición de conocimientos basados en los datos de su propia operación. Esto les permite realizar tareas como identificar fraudes, realizar perfiles de clientes con comportamientos similares, predecir cuales de los clientes están en mayor riesgo de abandonar una compañía por la competencia, identificar el mejor conjunto de productos a recomendar a un cliente, etc.

Este curso introductorio presenta la analítica de datos, ofreciendo técnicas de análisis de la información, con orígenes en la estadística y la inteligencia artificial, que permiten extraer conocimientos, que yacen profundamente en los datos. Así mismo, ofrece los elementos necesarios para que el estudiante se apropie del análisis de datos, formule preguntas correctas de investigación respecto a un fenómeno o situación, y ofrezca respuestas exploratorias utilizando técnicas de clasificación, regresión, segmentación y sistemas de recomendación.

Este curso es recomendado para estudiantes en proyecto de grado o jóvenes investigadores.

Objetivos

Generales

Al finalizar el curso, el estudiante estará en la capacidad de formular preguntas exploratorias a acerca de un comportamiento o fenómeno representado en un conjunto de datos. Para esto el estudiante podrá llevar a cabo los pasos necesarios para el robusto y correcto análisis y visualización de información, y para la generación de reportes o productos visuales derivados de dicho análisis.

Terminales

- Identificar los diferentes problemas que se pueden solucionar con la analítica, así como las técnicas correspondientes para encontrar una solución.
- Definir preguntas de investigación de analítica de datos
- Explorar y limpiar conjuntos de datos

- Crear modelos de analítica e interpretar sus resultados
- Evaluar y comparar modelos de analítica

Específicos

De formación académica:

Al finalizar cada unidad, el estudiante estará en capacidad de:

Unidad 1: Analítica, big data y metodología:

- Explicar que se entiende por analítica y big data
- Explicar las problemáticas que los grandes volúmenes de datos causan en las organizaciones
- Comprender el proceso global de la analítica de datos
- Describir las fases de las metodologías CRISP-DM y ASUM para proyectos de analítica
- Definir una pregunta de investigación

Unidad 2: Exploración de los datos

- Seleccionar un conjunto de datos de interés.
- Explorar conjuntos de datos de manera eficiente.
- Refinar la pregunta de investigación de acuerdo a evidencia temprana.
- Limpiar un conjunto de datos.

Unidad 3: Pruebas de hipótesis

- Identificar preguntas que se deban responder a partir de pruebas de hipótesis
- Calcular intervalos de confianza sobre muestras de datos
- Aplicar pruebas de hipótesis de relación entre una variables categóricas y numéricas
- Interpretar los resultados de las pruebas de hipótesis

Unidad 4: Modelos de analítica basados en aprendizaje supervisado

- Identificar preguntas que impliquen soluciones a partir de modelos de aprendizaje supervisado
- Aprender modelos de clasificación y regresión
- Interpretar los resultados de los modelos de clasificación y regresión
- Evaluar y comparar modelos de clasificación y regresión
- Seleccionar los mejores predictores
- Implementar protocolos de evaluación con el fin de evitar el sobre-aprendizaje

Unidad 5: Modelos de analítica basados en aprendizaje no supervisado

- Identificar preguntas que impliquen soluciones a partir de modelos de aprendizaje no supervisado
- Aplicar técnicas de reducción de dimensionalidad
- Segmentar conjuntos de datos utilizando modelos de clustering
- Evaluar segmentaciones obtenidas a partir de diferentes modelos de clustering

Unidad 6: Sistemas de recomendación

- Aplicar técnicas de sistemas de recomendación basadas en reglas de asociación
- Aplicar técnicas de sistemas de recomendación basadas en contenido
- Aplicar técnicas de sistemas de recomendación basadas en filtro colaborativo

De formación en valores y competencias

Contribuir a la formación de valores y habilidades tales como:

- La perseverancia y la autonomía, a través del desarrollo de los ejercicios que entregan los profesores.

- La tolerancia, mediante el respeto ante la participación de los compañeros y el profesor en clase (que muchas veces no coinciden con la propia opinión).
- La capacidad de análisis y conceptualización, a través de la identificación y modelado de los elementos necesarios para resolver un problema planteado.
- La capacidad de síntesis cuando deben elaborar una solución final utilizando elementos ya desarrollados.
- La capacidad de manejo de la información, mediante la limpieza de un conjunto de datos y la aplicación de algoritmos que ayuden en la solución de una pregunta planteada.
- La capacidad de trabajo bajo presión, ya que se cuenta con tiempos limitados para algunas actividades.

Metodología

Es un curso teórico práctico en el que se introducen conceptos fundamentales y luego se realizan proyectos cortos de aplicación. Todos los conceptos se aplicarán en un proyecto final que se desarrollará a lo largo del semestre.

Los estudiantes deberán preparar, antes de la clase, los temas asignados por el profesor, para lo cual:

- Estudiarán el material asignado en la programación del curso.
- Contestarán las preguntas que contiene el material, así como las preguntas adicionales que el profesor entregue.
- Resolverán los ejercicios propuestos en el material, así como los ejercicios adicionales que se le entreguen.
- Asistirán a las sesiones de monitoría o tutoría como parte del proceso de aprendizaje, aprovechando estos espacios para afianzar los temas vistos en clase.
- La preparación previa de los temas debe ser una labor constante, el estudiante deberá traer preguntas, dudas y comentarios a cada una de las sesiones. Debe estar listo para las pruebas o ejercicios que se planteen en clase.

Durante la clase, el estudiante deberá:

- Plantear las dudas que le quedaron durante el proceso de estudio del tema a tratar.
- Participar en las actividades de revisión y consolidación de conceptos que proponga el profesor.
- Trabajar en la solución de los problemas de aplicación que se propongan.

Después de la clase:

- Establecer las relaciones entre los temas tratados en la clase y el conocimiento previamente adquirido.
- Resolver los ejercicios de aplicación del tema, que tienen un nivel de complejidad mayor al de los ejercicios que resolvió previamente.
- Adicionalmente, si ya se ha planteado el proyecto de la unidad debe trabajar individual o grupalmente en el mismo, según el caso: identificar las relaciones entre los conceptos recién aprendidos y los conceptos previamente trabajados.

Este curso tiene dos componentes, con igual importancia:

- Componente Teórica: Consiste en la discusión de los diferentes temas y el hacer ejercicios que permitan ponerlos en práctica, esto en el salón de clase.
- Componente Práctica: Consiste en la resolución de ejercicios y su implementación en la sala de cómputo.

Evaluación

La evaluación del curso comprende los temas en sus aspectos tanto teóricos como prácticos.

El proyecto final se trabajará en grupo y su sustentación será tanto individual como grupal.

Forma de Evaluación	Porcentaje	Cuándo se realiza
Parcial 1	20%	Sesión 4
Parcial 2	20%	Sesión 8
Entrega preliminar del proyecto final	15%	Sesión 12
Parcial 3	20%	Sesión 15
Entrega y sustentación del proyecto final	25%	Sesión 16
Total	100%	

De conformidad con lo establecido en el reglamento estudiantil en cuanto a la evaluación, para que un estudiante apruebe el curso deberá tener un promedio aritmético de 3.0 o superior, calculado con las notas individuales que haya obtenido en las tareas y comprobaciones de lectura. Si cumple lo anterior, la nota definitiva corresponderá al cálculo de las notas con la ponderación del cuadro anterior. En caso contrario, la nota definitiva corresponderá a la del promedio ponderado con las notas individuales que haya obtenido, discriminadas según la tabla anterior.

Bibliografía

- An Introduction to Statistical Learning with Applications in R. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. Springer. 2009. Gratis on-line (<http://www-bcf.usc.edu/~gareth/ISL/>)
- R for Data Science: Import, Tidy, Transform, Visualize and Model Data. Hadley Wickham & Garret Grolemund. O'Reilly, 2017
- Swirl. Learn R in R. <http://swirlstats.com/>
- Software de referencia
 - R <http://www.r-project.org/>
 - R Studio <http://www.rstudio.com/>
 - R Markdown <http://rmarkdown.rstudio.com/>
 - Weka