



Análisis de Datos

2 teóricos, 2 prácticos



A. Información del profesor

Nombre del profesor

Stanly Bolaños

Correo electrónico

dsbolanosa@correo.url.edu.gt

Campus o sede

Campus Central

Horario

Lunes y miércoles.
19:30-21:00



B. Información general

Descripción

Este curso busca introducir a los estudiantes a los conceptos del análisis de información, el cual representa hoy en día uno de los insumos más importantes para las empresas al momento de tomar decisiones y definir planes estratégicos. Mucho de dicho análisis se basa en realizar lo que se conoce como un análisis "**descriptivo**" de la información histórica, en el cual se procesa la información con el objetivo de definir y evaluar qué es lo que ha pasado anteriormente, y aunque el mismo es de gran utilidad para mostrar cómo ha estado el desempeño pasado, es muchas veces el análisis "**predictivo**" en el que las empresas invierten para poder determinar en el futuro cuáles serán las mejores decisiones a tomar. Por último, estudiaremos también como a través de las técnicas avanzadas de ciencias de datos y ramas de la inteligencia artificial como el "machine learning" podemos realizar el análisis "**prescriptivo**" que es cuando no solo somos capaces de predecir (en función del pasado) sino también automatizar el plan de acción que se debe de tomar.

Modalidad

Virtual



C. Malla curricular

COMPETENCIAS GENÉRICAS



El egresado landivariano se identifica por:

Pensamiento lógico, reflexivo y analógico	Pensamiento crítico	Resolución de problemas
Habilidades de investigación	Uso de TIC y gestión de la información	Comunicación efectiva, escrita y oral
Comprensión lectora	Compromiso ético y ciudadanía	Liderazgo constructivo
Aprecio y respeto por la diversidad e interculturalidad	Creatividad	

COMPETENCIAS ESPECÍFICAS (propias del curso)

Competencia 1

Introducción al Análisis de Datos

Competencia 2

Business Intelligence

Competencia 3

Machine Learning

Competencia 4

Inducción al Análisis Avanzado de Datos



METODOLOGÍA

El curso se desarrollará intercalando clases magistrales para la exposición de conceptos nuevos y clases participativas, en las que se asume y comprueba que el estudiante realiza las lecturas previas, tareas o ejercicios dejados para realizar fuera de clase, previo al inicio de un nuevo día de clase. Además de ejercicios prácticos relacionados a temas tratados en clase o en días anteriores. Laboratorios en los cuales se profundizará en ejercicios que afiancen los conceptos dados y la exploración del manejador de base de datos seleccionado.



PROGRAMACIÓN

Programación Indicador de logro 1		
Introducción al Análisis de Datos		
Metodología: Clases magistrales, laboratorios y clases interactivas.		
Saber conceptual (temas)	Saber procedimental (habilidades, destrezas, procesos)	Saber actitudinal (conductas observables)
Introducción a Análisis y Big Data Ciclo de Vida del Análisis de Datos	<ul style="list-style-type: none"> - Conceptos Generales - Familiarizarse con las perspectivas de análisis Descriptivo, prescriptivo, predictivo - CRISP-DM 	Entendimiento general del ciclo de vida de los proyectos de análisis de datos y definición conceptual.



Programación Indicador de logro 2

Business Intelligence

Metodología: Clases magistrales, laboratorios y clases interactivas.

Saber conceptual (temas)	Saber procedimental (habilidades, destrezas, procesos)	Saber actitudinal (conductas observables)
Business Intelligence - Dimensiones, Facts y Measures Business Intelligence - DWH, SCD y ETL Business Intelligence - Cubos y Herramientas OLAP Visualización de Datos - Reportes y Dashboards Visualización de Datos – Tableau / Power BI	<ul style="list-style-type: none"> - Diferenciación entre dimensiones, medidas, tablas de dimensión y tablas de hechos. - Comprender el ciclo de extracción transformación y carga de datos. - Utilizar el proceso ETL para el llenado de información del Data Warehouse. - Creación de modelos empresariales para la extracción de información. - Aprender a utilizar herramientas de visualización para presentación de resultados. - Identificar las diferentes herramientas de visualización para la correcta elección en los proyectos. - Aprender los conceptos del paradigma “Self Service” en BI para la distribución de información empresarial. 	Implementar una solución Business Intelligence para proporcionar modelos empresariales



Programación Indicador de logro 3

Machine Learning

Metodología: Clases magistrales, laboratorios y clases interactivas.

Saber conceptual (temas)	Saber procedimental (habilidades, destrezas, procesos)	Saber actitudinal (conductas observables)
Análisis de datos Usando R - Introducción a R y lenguajes estadísticos Análisis de datos Usando R - Estadística para la construcción de modelos Análisis de datos Usando R - Análisis de Hipotesis Análisis avanzado de datos - Clustering Análisis avanzado de datos - Reglas de Asociación Análisis avanzado de datos - Regresión Análisis avanzado de datos - Clasificación Análisis avanzado de datos - Machine Learning as a Services y AutoML	<ul style="list-style-type: none"> - Aprender a desplegar un entorno R Studio para el desarrollo de proyectos. - Instalación y despliegue de paquetes y librerías en R. - Elaboración de modelos estadísticos a partir de funciones en R. - Graficar y visualizar datos utilizando librerías (ggplot2) - Diferencias modelos de aprendizaje supervisado y no supervisado aplicando las distintas técnicas de modelado. - Aprender nuevas herramientas de machine learning para la automatización de los algoritmos de aprendizaje. - Implementar algoritmos de R en visualizadores trabajados (PowerBI, Tableau) 	Deferencias los beneficios de los distintos modelos de Machine Learning según el caso de negocio requerido.



Programación Indicador de logro 4

Análisis avanzado de datos

Metodología: Clases magistrales, laboratorios y clases interactivas.

Saber conceptual (temas)	Saber procedimental (habilidades, destrezas, procesos)	Saber actitudinal (conductas observables)
- MapReduce y Hadoop	- Identificar la arquitectura Hadoop aprovechando la eficiencia proporcionada por MapReduce (HDFS)	Conocer herramientas para el manejo de BigData.



EVALUACIÓN

a. Estrategias de evaluación sumativa

Estrategias	Puntaje
Pruebas cortas	10
Laboratorios y hojas de trabajo	20
Evaluación 1	15
Evaluación 2	15
Proyecto 1	10
Proyecto 2	10
Evaluación Final	20
TOTAL	100

b. Estrategias de evaluación formativa

Técnicas formativas	Utilizadas en el curso	Procedimiento
Casos	x	Casos para trabajar de forma individual o colectiva
Padlet colaborativo	x	Se usará para compartir resultados de algunos ejercicios y recibir retroalimentación
Exámenes cortos sumativos	x	Exámenes con punteo de zona
Trabajos en pequeños grupos para resolver dudas	x	Se espera que los estudiantes se apoyen en el aprendizaje trabajando colaborativamente
Laboratorios	x	Ejercicios en actividad de laboratorio que se realiza con periodo doble



CALENDARIO DE REFERENCIA POR TEMAS



Fecha	Tema	Actividad de evaluación
Semana 1	Introducción a Análisis y Big Data	Tarea – Demos de Big Data
Semana 2	Ciclo de Vida del Análisis de Datos	Tarea – Ciclo de vida en caso de estudio corto
Semana 3	Business Intelligence – DWH, SCD y ETL	Hoja de Trabajo # 1 – Queries y Modelo DWH
Semana 4	Business Intelligence – Cubos y Herramientas OLAP	Hoja de Trabajo # 2 – SSIS Evaluación Corta # 1
Semana 5	Visualización de Datos - Reportes y Dashboards	Hoja de Trabajo # 3 - SCD y Cubo OLAP
Semana 6	Visualización de Datos – Tableau / Power BI	Evaluación Parcial # 1
Semana 7	Análisis de datos Usando R - Introducción a R y lenguajes estadísticos	Hoja de Trabajo # 4 - Reportes y Dashboards
Semana 8	Análisis de datos Usando R - Estadística para la construcción de modelos	Evaluación Corta # 2 Entrega Proyecto # 1
Semana 9	Análisis de datos Usando R - Análisis de Hipotesis	Hoja de Trabajo # 6 - Exploración e Hipótesis
Semana 10	Análisis avanzado de datos - Clustering	Evaluación Parcial # 2
Semana 11	Análisis avanzado de datos - Reglas de Asociación	Hoja de Trabajo # 7 - Clustering y Asociación
Semana 12	Análisis avanzado de datos - Regresión	Hoja de Trabajo # 8 Regresión Evaluación
Semana 13	Análisis avanzado de datos - Clasificación	Hoja de Trabajo # 9 Clasificación Evaluación Corta # 3
Semana 14	Análisis avanzado de datos – Machine Learning as a Services y AutoML	Entrega Proyecto # 2
Semana 15	Análisis avanzado de datos - MapReduce y Hadoop	Hoja de Trabajo # 10 Hadoop Evaluación Final

REFERENCIAS BIBLIOGRÁFICAS

O'Neil, C. (2013). Doing Data Science: Straight Talk from the Frontline. Editorial O'Reilly

Kimball, R. (2008) The Data Warehouse Lifecycle Toolkit, Second Edition. Editorial Wiley