Diagrama

Descripción generada automáticamente

Semana 1

|  |  |  |
| --- | --- | --- |
| **Tema** | **Síntesis de los contenidos abordados en la sesión** | **Lecciones aprendidas** |
| **Instalación y configuración de los entornos** | Proceder con la instalación de Anaconda, Python y Jupyter Notebook, el cual será la suite para los proyectos de ciencia de datos.  Para el trabajo colaborativo y repositorio de trabajos es indispensable el uso de GitHub. | * Disponer de las herramientas necesarias para trabajar con datos. * El entorno en que se trabajará para crear el algoritmo es Anaconda. * El lenguaje para el desarrollo del código será Python. * Para trabajar este cuatrimestre se utilizará GitHub. |
| **Desarrollo de actividades y lecturas complementarias** | Anaconda Navigator es un IDE creado para facilitar y concentrar las principales herramientas para la ciencia de datos a nivel mundial, dentro de las cuales se encuentran Jupyter Lab, Jupyter Notebook, Qtconsole, Spyder, Vscode, Glueviz, Orange3 y Rstudio.  Jupyter Notebook es un cuaderno de trabajo en línea mediante el cual se pueden crear anotaciones de diferentes tipos. Es la mas utilizada para proyectos de análisis de datos con Python.  GitHub es un servicio basado en la nube que aloja un sistema de control de versiones, esto permite a los desarrolladores colaborar y realizar cambios en proyectos compartidos, a la vez mantienen un seguimiento detallado de su progreso. La interfaz de usuario es fácil de utilizar, por lo que la hace accesible para personas con poco o ningún conocimiento técnico. | |
| **Evidencias** | Ambiente de Anaconda posterior a la instalación y video sobre introducción a Jupyter Notebook. | |

Semana 2

|  |  |  |
| --- | --- | --- |
| **Tema** | **Síntesis de los contenidos abordados en la sesión** | **Lecciones aprendidas** |
| **Repaso de los principales conceptos de programación y Python.** | Se realiza el desarrollo de un código para refrescar conceptos como variables, ciclos y otros.  Variables: valor que puede cambiar o contenedor que puede cambiar su valor, es un elemento que puede estar cambiante en el código, y puede contener un dato conocido o desconocido. Dentro de los tipos de variables se encuentran las no numéricas (carácter o lógicos) y las numéricas (decimales o enteros. Finalmente se realizó un resumen con las sanas practicas a la hora de crear variables.  Ciclos: El “For” es utilizado cuando sabemos la cantidad de interacciones y el “While” cuando desconocemos esa cantidad.  Funciones: Son microprogramas que buscan resolver un problema focal, sirven para no estar repitiendo el código. | * Repaso del tema de variables (que son, tipos de datos y variables, y sanas practicas), ciclos for y while; y funciones, mediante un ejercicio práctico. |
| **Desarrollo de actividades y lecturas complementarias** | Metodología CRISP-DM fue la primera metodología en existir y constituye el inicio para otras metodologías, siendo de hasta hoy una de las de mayor uso. Es flexible y se puede personalizar fácilmente, permitiendo que se adapte a las necesidades concretas.  Se compone de 6 fases: Comprensión del negocio, enfoque analítico, preparación de los datos (requisitos de datos, recopilación y comprensión de los datos), modelado, evaluación e implementación (retroalimentación).  Metodología BADIR (proviene del libro Behind Every Good Decision) es utilizada como Framework para gestionar correctamente los procesos de análisis de datos. Se encuentra estructurada en 5 pasos: Business question, Analysis plan, Data collection, Insights and Recommendations. | |
| **Evidencias** | Resolución de ejercicio práctico sobre realización de un sistema de baja complejidad donde se identifiquen los números primos o compuestos en un rango dado por el usuario.  Texto  Descripción generada automáticamente  Repaso de variables, ciclos y funciones | |

Semana 3

|  |  |  |
| --- | --- | --- |
| **Tema** | **Síntesis de los contenidos abordados en la sesión** | **Lecciones aprendidas** |
| **Metodología CRISP-DM y BADIR**  **HTML & Jupyter Notebook** | Conversación sobre las 5 etapas de la metodología para el análisis de datos.  Metodología CRISP-DM: (Cross Industry Standard Process for Data Mining) es un método probado para orientar los trabajos de minería de datos.  Como metodología, incluye descripciones de las fases normales de un proyecto, tareas necesarias en cada fase y una explicación de las relaciones entre las tareas.  Como modelo de proceso, ofrece un resumen del ciclo vital de minería de datos.  La metodología BADIR brinda una perspectiva diferente sobre cómo abordar los procesos a realizar dentro de los proyectos de análisis de datos. Consta de los siguientes pasos:   1. Entender el negocio: Lograr comprender la pregunta a responder. 2. Comprender los datos- recolección de datos: Como se tiene que interpretar esos datos. Las empresas tienen Datawarehouse, archivos Excel, Txt, documentos físicos, etc. 3. Preparación de los datos: Como manejar los datos para el análisis y seleccionar las variables correctas. 4. Modelado de algoritmos: regresión, clasificación y clusterizacion. 5. Evaluación de resultados: Evalúa el rendimiento de los algoritmos. 6. Implementación- Puesta en marcha.   HTML (Hyper Text Markup Language) es un lenguaje marcado para la descripción de documentos web (páginas web). Cada etiqueta HTML describe diferentes contenidos en el documento.   * En cuanto a la estructura, está conformado por bloques que abren con <xx> y cierran con </xxx>. * Los títulos se definen con etiquetas de h1 al h6. El formato es <h1>xxx </h1>. * Los párrafos se definen con la etiqueta <p>xxx</p>. * Los vínculos <a href=>link<a/>. * Para insertar imágenes <img src=”imagen.jpg”, las cuales deben estar guardadas en la misma carpeta donde se está trabajando. * Para agregar colores <h1 style=”color: #0000FF”> xxxx </h1>, es decir, se debe utilizar el código hexadecimal de los colores. | * Se deben anonimizar los datos. Cuando se trabajan con datos personalizados hay que tener presente las leyes que regulan esta información. * Se busca una efectividad del algoritmo del 80 al 95%. * La metodología BADIR para explicar los proyectos de análisis de datos es perfecta ya que es de más fácil comprensión para los que no conocen el proceso. * Con la metodología CRISP-DM se abordan todos los aspectos o fases indispensables para un correcto proceso de análisis de datos. * Importancia de las habilidades informáticas, matemáticas- estadísticas y la experiencia del negocio para poder crear la ciencia de datos. * Conocimiento en el uso básico del lenguaje HTML como estructura, títulos, párrafos, vínculos insertar imágenes, cambios de color dentro de la plataforma de Jupyter Notebook. |
| **Desarrollo de actividades y lecturas complementarias** | Estadística descriptiva comprende un conjunto de técnicas para un análisis preliminar de los datos que permite familiarizarse con ellos y descubrir sus principales características mediante operaciones exploratorias; y trasmitir el conocimiento adquirido sobre los datos a terceros.  Esta relacionada a el EDA (Exploratory data análisis) que extiende la exploración grafica de los datos hasta incluir la fase de análisis.  Abarca temas como los resúmenes números que contienen los usos de funciones estadísticas, las tablas de datos, visualizaciones de datos, análisis exploratorios de datos (EDA), modelos como herramientas descriptivas y consideraciones relevantes como los outliers, los nulos y las transformaciones de variables.  Aplicación del EDA en proyectos con Python se incluye una guía práctica para el análisis exploratorio de datos con Pandas, en cuanto a temas como importar pandas, crear un DataFrame, visualizar un DataFrame, leer y guardar archivos CSV, realizar resúmenes estadísticos, etc. | |
| **Evidencias** | Explicación sobre la metodología CRISP-DM y BADIR      Explicación sobre la utilización de HTML en Jupyter Notebook | |