Universidad Castro Carazo

Jimena Badilla

Portafolio de evidencias

Programación

# Semana 2

|  |  |  |
| --- | --- | --- |
| ¿Qué vimos? | ¿Qué aprendí? |  |
| Repaso de Variables  CRISP-DM  BADIR  Entender el negocio  Comprender datos  Preparación de datos  Modeling | Las variables son un espacio que contiene dato conocido o desconocido.  Importante: Puede variar su valor en cualquier momento.  Metodología probada para orientar trabajos de minería de datos.  6 fases del ciclo: Business understanding, Data Understanding, Data preparation, Modeling, Evaluation and Deployment.  Business questions, Analysis plan, Data collection, Insights and Recommendations.  Comprender la pregunta a responder.  “Los datos no hablan, pero si responden.  Datawarehouse, Excel, archivos  Como manejar los datos para el análisis, seleccionar variables correctas para ML.  Regresión  Clasificación  Clusterizacion |  |

## Evidencias

A picture containing diagram

Description automatically generated

# Semana 4

|  |  |  |
| --- | --- | --- |
| ¿Qué vimos? | ¿Qué aprendí? |  |
| HTML Y JUPITER  Carga de Set de Datos para Análisis | La importancia de traer imágenes me evita estar abriendo archivos. Únicamente se hace un scroll down a la imagen para buscar los datos.  Como color títulos en HTML, color, párrafos etc.  Como cargar las imágenes.  Diferentes librerías y sus usos como:  Pandas – Datagrame  Numpy - Manejo de estructura de datos  Matplotlib – gráficos |  |

Evidencia de Actividad

Graphical user interface, text, application, email

Description automatically generated

# Semana 5

|  |  |  |
| --- | --- | --- |
| ¿Qué vimos? | ¿Qué aprendí? |  |
| Manejo de valores faltantes  Iloc:  Loc: | Todo lo que sea valores numéricos se pueden sustituir por la media.  La media se calcula con decimales en Python.    Se puede poner por index, o ID que consume menos memoria..  Se puede escoger por nombre de columna |  |

Evidencia de Actividad

Graphical user interface

Description automatically generated with medium confidence

# Semana 6

|  |  |  |
| --- | --- | --- |
| ¿Qué vimos? | ¿Qué aprendí? |  |
| Manejo de valores faltantes  Iloc:  Loc: | Todo lo que sea valores numéricos se pueden sustituir por la media.  La media se calcula con decimales en Python.    Se puede poner por index, o ID que consume menos memoria..  Se puede escoger por nombre de columna |  |

# Semana 7

|  |  |  |
| --- | --- | --- |
| ¿Qué vimos? | ¿Qué aprendí? |  |
| Bining de Datos con diferentes librerías.  Pandas:  Matplotlib:  Dato categórico:  Observaciones:  Scipy.stacts:  SKLEARN: | Cargar el DataFrame  Librería para cargar gráficos, PLT: Le da color a la gráfica.  Dato numérico que pasa a ser carácter. Ejm: 4= menor  10= Mayor      Es equivalente a registros  Librería de intervalos, variables etc.  Utilizada para machine learning. |  |