Fundación Universitaria Konrad Lorenz.
Big Data Electiva-I. Parcial-1. Diabetes Clinical Data.
Estudiante: David Gutierrez Chaves Cod: 506222728.

Parcial #1:

Instrucciones:
• Fecha de publicación: 08 de Marzo de 2025
• Fecha de entrega: 15 de Marzo 2025
• Medio de entrega: Sustentación en clase.

Conjunto de datos: Contiene datos sobre datos clínicos de pacientes.
https://www.kaggle.com/datasets/ziya07/diabetes-clinical-dataset100k-rows

Ejercicio:
° El Dataset entregado presenta los datos de pacientes que presentan afectaciones de diabetes. La idea es analizar cuáles son las causas para que un paciente tenga afectaciones de salud por esta enfermedad. Se debe realizar lo siguiente:
° Preprocesamiento de los datos: limpieza, revisión de datos faltantes/anómalos, reducción de datos (seleccionar características relevantes) y transformaciones.
° Análisis exploratorio (EDA) de los datos para responder al problema: ¿Qué características hacen que se diagnostique a una persona con diabetes?
° Entrenamiento de una red neuronal.

Entregables:
· Análisis exploratorio en notebook (10 puntos).
· Preprocesamiento de los datos (15 puntos).
· Entrenamiento del Modelo de Machine Learning con redes neuronales (10 puntos).
· Evaluación del modelo 75% Accuracy (10 puntos).

Tema: Diabetes Clinical Data
Diabetes Clinical Data: https://github.com/dg2c4/Diabetes-Clinical-Data

Problema: ¿Qué características hacen que se diagnostique a una persona con diabetes?
La diabetes es una enfermedad metabólica compleja que requiere un análisis exhaustivo de múltiples características clínicas y biométricas para su diagnóstico preciso. El siguiente análisis detalla las características principales que determinan el diagnóstico de la diabetes y cómo estas se procesan mediante técnicas de big data y aprendizaje automático.

Objetivo:
Desarrollar un análisis completo que permita extraer datos sobre los pacientes que presentan afectaciones de diabetes, utilizando herramientas de modelos machine learning y el algoritmo Kmeans, para visualizar los resultados de manera efectiva. El trabajo se divide en tres componentes principales:

Análisis Exploratorio:
- Formulación de preguntas de investigación.
- Creación de hipótesis basadas en un problema específico.
- Análisis inicial del conjunto de datos seleccionado.

Fundación Universitaria Konrad Lorenz.
Big Data Electiva-I. Parcial-1. Diabetes Clinical Data.
Estudiante: David Gutierrez Chaves Cod: 506222728.

Preprocesamiento de Datos:
- Limpieza de datos.
- Transformación de variables.
- Reducción de datos cuando sea necesario.
- Discretización de datos según corresponda.

Modelado y Evaluación:
- Implementación de modelo de regresión lineal o logística.
- Selección y cálculo de métricas de evaluación.
- Análisis del rendimiento del modelo.

Systems engineering:
User: David Gutierrez Chaves
Code: 506222728
Subject: Big Data Electiva-I
Institution: Fundación Universitaria Konrad Lorenz
Institutional email: david.gutierrec@konradlorenz.edu.co

Created By: https://github.com/dg2c4