Fundación Universitaria Konrad Lorenz.
Big Data Electiva-I. Taller #1. Apple Quality Machine Learning Model.
Estudiante: David Gutierrez Chaves Cod: 506222728.

Taller #1:

Instrucciones:
• Fecha de publicación: 08 de febrero de 2025
• Fecha de entrega: 01 de marzo de 2025.
• Medio de entrega: Sustentación en clase.

Conjunto de datos: Contiene información sobre la calidad de las manzanas.
https://www.kaggle.com/datasets/nelgiriyewithana/apple-quality

Ejercicio:

A. Realizar un análisis exploratorio al conjunto de datos entregado, para ello debe crear diferentes preguntas e hipótesis a resolver en los datos partiendo de un problema que cada grupo debe plantear.
B. Después de realizar el análisis exploratorio, debe hacer el preprocesamiento de los datos según como considere: limpieza, transformación, reducción de datos o discretización de los datos. Para esto, puede usar pandas.
C. Luego de hacer el preprocesamiento y tener el conjunto de datos final, debe crear un modelo de Machine Learning de regresión lineal, puede usar scikit learn. Este modelo puede ser de predicción o regresión logística para clasificación.
D. Debe buscar la forma de evaluar si el modelo está bien o no con métricas que permitan determinar si el modelo aprende o no.

Entregables:
Análisis exploratorio en notebook (10 puntos).
Preprocesamiento de los datos (15 puntos).
Entrenamiento del Modelo de Machine Learning (15 puntos).
Evaluación del modelo (10 puntos).

Tema: Modelo Machine Learning.
Apple-Quality-Machine-Learning-Model:
https://github.com/dg2c4/Apple-Quality-Machine-Learning-Model

Objetivo:
Desarrollar un modelo que aprenda efectivamente de los datos y pueda realizar predicciones o clasificaciones precisas, utilizando herramientas como pandas para el manejo de datos y scikit-learn para el modelado.

Desarrollo:
Este trabajo de Machine Learning se centra en el desarrollo de un proyecto completo de análisis de datos, desde la exploración inicial hasta la evaluación de un modelo predictivo. El proceso se divide en tres fases principales:

Fundación Universitaria Konrad Lorenz.
Big Data Electiva-I. Taller #1. Apple Quality Machine Learning Model.
Estudiante: David Gutierrez Chaves Cod: 506222728.

Análisis Exploratorio:
- Identificación de preguntas de investigación.
. Formulación de hipótesis.
- Análisis inicial del conjunto de datos.

Preprocesamiento de Datos:
- Limpieza de datos.
- Transformación de variables.
- Reducción de dimensionalidad.
- Discretización cuando sea necesaria.

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