

Documento sobre el dataset código empleado para cargarlo en el sistema

Datos sobre el manejo del dataset

Para operación del usuario la data fue cargada en one drive para llamar en google colab mediante el siguiente código.

```
# Celda 3: Descarga del Dataset (condicional)
file_path = "LSWMD.pkl"
file_id = "1JGJ_a7c5Cs4cAGh7NGElfGfhIm90-WUj"
url = f"https://drive.google.com/uc?id={file_id}"
```

```
if not os.path.exists(file_path):
    print("Downloading dataset...")
    gdown.download(url, file_path, quiet=False)
print("Dataset found!")
```

Con esto se cargó el archivo

```
import pandas as pd
# Ruta del archivo .pkl
file_path = "LSWMD.pkl"
# Cargar el archivo .pkl
try:
    df = pd.read_pickle(file_path)
    print("Archivo cargado correctamente.")
except Exception as e:
    print("Error al cargar el archivo:", e)
```

Se mostró la data del dataframe

```
# Verificar la información del DataFrame
df.info()
```

```
# Mostrar las primeras filas del DataFrame
print(df.head())
# Verificar las columnas del DataFrame
print("Columnas del DataFrame:", df.columns)
# Verificar estadísticas descriptivas
print(df.describe())
```

el resultado obtenido fue este

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 811457 entries, 0 to 811456
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   waferMap     811457 non-null object
```

```

1 dieSize      811457 non-null float64
2 lotName      811457 non-null object
3 waferIndex   811457 non-null float64
4 trianTestLabel 811457 non-null object
5 failureType  811457 non-null object

```

```
dtypes: float64(2), object(4)
```

```
memory usage: 37.1+ MB
```

```

           waferMap dieSize lotName \
0 [[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,... 1683.0 lot1
1 [[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,... 1683.0 lot1
2 [[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,... 1683.0 lot1
3 [[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,... 1683.0 lot1
4 [[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,... 1683.0 lot1

```

```

waferIndex trianTestLabel failureType
0    1.0 [[Training]] [[none]]
1    2.0 [[Training]] [[none]]
2    3.0 [[Training]] [[none]]
3    4.0 [[Training]] [[none]]
4    5.0 [[Training]] [[none]]

```

```

Columnas del DataFrame: Index(['waferMap', 'dieSize', 'lotName', 'waferIndex', 'trianTestLabel',
                                'failureType'],
                                dtype='object')

```

```

           dieSize waferIndex
count 811457.000000 811457.000000
mean  1840.998585   12.961207
std    2254.987374    7.207489
min     3.000000    1.000000
25%    710.000000    7.000000
50%    953.000000   13.000000
75%   1902.000000   19.000000
max   48099.000000   25.000000

```