

procesamiento2

June 9, 2025

1 Conversión del dataset en csv a arff

1.1 Importación de los datos

En primer lugar, es necesario importar el dataset procesado y en formato csv.

```
[1]: # Instalar la librería liac-arff para manejar archivos ARFF
%pip install liac-arff
```

```
Collecting liac-arff
  Downloading liac-arff-2.5.0.tar.gz (13 kB)
  Installing build dependencies: started
  Installing build dependencies: finished with status 'done'
  Getting requirements to build wheel: started
  Getting requirements to build wheel: finished with status 'done'
  Preparing metadata (pyproject.toml): started
  Preparing metadata (pyproject.toml): finished with status 'done'
Building wheels for collected packages: liac-arff
  Building wheel for liac-arff (pyproject.toml): started
  Building wheel for liac-arff (pyproject.toml): finished with status 'done'
  Created wheel for liac-arff: filename=liac_arff-2.5.0-py3-none-any.whl
  size=11782
  sha256=c15e8fa9bc7d6dff5e2b1b019b0cf002594229816980bc4e2ea812de6b76d806
  Stored in directory: c:\users\carolina\appdata\local\pip\cache\wheels\a9\ac\cf
  \c2919807a5c623926d217c0a18eb5b457e5c19d242c3b5963a
Successfully built liac-arff
Installing collected packages: liac-arff
Successfully installed liac-arff-2.5.0
Note: you may need to restart the kernel to use updated packages.
```

```
[2]: # Importar las librerías necesarias
import pandas as pd
import arff
```

```
[4]: # Cargar el dataset CSV
df = pd.read_csv("../Data/DP_notas_categorizadas.csv")

# Mostrar las primeras filas del DataFrame
print(df.head())
```

	cod_escuela	cod_plan	cod_asignatura	categoria_nota
0	1	2009	201003	Aceptable
1	1	2009	201101	Aceptable
2	1	2009	201101	Reprobado
3	1	2009	201101	Aceptable
4	1	2009	201101	Baja

1.2 Preparación de los datos

En segundo lugar, se preparan los datos para garantizar la compatibilidad entre csv y arff

```
[5]: # Preparar los atributos para ARFF
attributes = [(col, df[col].unique().tolist() if df[col].dtype == 'object' else
↳ 'REAL') for col in df.columns]

# Crear estructura ARFF
arff_data = {
    'description': 'Dataset de notas categorizadas de universitarios de la
↳ FISI-UNMSM',
    'relation': 'notas_categorizadas',
    'attributes': attributes,
    'data': df.values.tolist()
}
```

1.3 Exportación del archivo ARFF

Finalmente, se exporta el dataset como archivo ARFF, lo cual facilita su tratamiento con WEKA.

```
[6]: # Guardar como archivo ARFF
with open("../Data/DP_notas_categorizadas.arff", "w") as f:
    arff.dump(arff_data, f)

print(" Archivo ARFF generado con éxito.")
```

Archivo ARFF generado con éxito.