



## **APRENDIZAJE AUTOMÁTICO**

### **Evaluación Parcial**

#### **Etaapa 2: Descripción del Dataset y Origen**

**Profesor Lic. Martin Mirabete**

**Alumna: Demari Monica Valeria**

- **DATASET**

Para el desarrollo del proyecto de Aprendizaje Automático utilizaré el archivo Excel llamado "dataset\_jcyd\_ok.xlsx"

- **INFORME SOBRE EL ORIGEN DEL DATASET**

Para llevar cabo el proyecto de optimizar la asignación y clasificación de docentes en la educación secundaria, presenté nota a la Secretaría de Gestión Educativa del ministerio de Educación con el fin poder obtener información que posee la Junta de Clasificación y disciplina de Educación Secundaria. El set de datos solicitado contendría información relacionada a los títulos de aspirantes inscriptos que se registran en la ciudad de Río Grande. Y, como respuesta a lo requerido, la Sra. secretaria de Gestión Educativa Prof. Silvina E. Solohaga, indicó la autorización a mi solicitud, siempre y cuando se resguarde la información individual y particular de los docentes.

Posteriormente, el equipo informático, con la revisión de los miembros de junta, me otorgaron el el archivo Excel llamado "dataset\_jcyd\_ok.xlsx", el cual contempla toda información hasta el último día de inscripción con fecha 30 de junio del 2023.

- **DESCRIPCIÓN DEL DATASET**

En el Jupyter "Descripción del dataset", se procedió cargar el dataset para dar inicio del proyecto de Aprendizaje Automático, y como información podemos destacar que este set está conformado por 52936 registros y 16 columnas.

El DataFrame contiene varias columnas y muestra el conteo de entradas no nulas (non-null) para cada una:



**TECNICATURA SUPERIOR EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL**

Orden	Nombre de columna	Entradas	Tipo de datos	Descripción
1.	idespacio	52,936	tipo entero (int64).	Clave de identidad del Espacio curricular
2.	Idtitulo	52,936	tipo objeto (string).	Clave de identidad del Título
3.	idtitulo_real	52,936	tipo entero (int64).	Clave de identidad del Título conformado entre los datos del título y la casa de estudio
4.	Idcasaestudio	52,936	tipo objeto (string)	Clave de identidad de la casa emisora del título
5.	Carácter	52,936	tipo objeto (string).	Nombre que se clasifican los títulos en el espacio curricular
6.	idNivel_espacio	52,935	tipo flotante (float64) - una entrada es nula.	Clave de identidad del Nivel del Espacio curricular o cargo
7.	desc_espacio	52,935	tipo objeto (string) - una entrada es nula.	Descripción del Cargo o Espacio curricular
8.	tipo_espacio	52,935	tipo objeto (string) - una entrada es nula.	Tipo de cargo o espacio curricular
9.	ciudad_espacio	52,935	tipo flotante (float64) - una entrada es nula.	La ciudad que se encuentra habilitado el cargo o espacio curricular
10.	resolucion_espacio	52,935	tipo objeto (string) - una entrada es nula.	Resolución del plan de estudio del título relacionado al cargo o espacio curricular
11.	titulo	52,811	tipo objeto (string) - 125 entradas son nulas.	Nombre del título
12.	resolucion	50,811	tipo objeto (string) - 2,125 entradas son nulas.	Resolución del plan de estudio del título
13.	nombre	52,695	tipo objeto (string) - 241 entradas son nulas.	nombre de la casa de estudio que emitió el título
14.	facultad	14,125	tipo objeto (string) - 38,811 entradas son nulas.	facultad de la casa de estudio que emitió el título
15.	provincia	50,367	tipo objeto (string) - 2,569 entradas son nulas.	provincia de la casa de estudio que emitió el título



## TECNICATURA SUPERIOR EN CIENCIA DE DATOS E INTELIGENCIA ARTIFICIAL

16.	ciudad	42,042	tipo objeto (string) - 10,894 entradas son nulas.	ciudad de la casa de estudio que emitió el título
-----	--------	--------	---	---

Por lo detallado en la tabla, el DataFrame tiene dos tipos de datos flotantes, dos enteros y doce objetos (cadenas de texto)

Las columnas que el tipo de datos es numérico se identifican como claves ID, por lo que es irrelevante proceder a hacer las estadísticas descriptivas de dichas columnas.

Las estadísticas descriptivas de las columnas categóricas presentan información sobre la cantidad de datos y la variedad de categorías en cada columna. De acuerdo a lo brindado en el Jupiter Notebook, puedo resaltar las siguientes observaciones más relevantes a mi investigación:

- Se distinguen tres categorías únicas, Docente "D", habilitante "H" y supletorio "S", siendo que la que mas prevalece es la habilitante con 20355 casos.
- En la columna descripción de espacios curriculares o cargos, se registran 468 descripciones únicas, destacando "Preceptor/a" como la más frecuente con 4648 casos de un total de 52935 registros.
- Con respecto a los títulos, figuran 2375 títulos únicos, con "PROFESOR/A DE EDUCACIÓN FÍSICA" destacando con 2153 repeticiones.