

UNIVERSIDAD NACIONAL DE HURLINGHAM  
INSTITUTO DE TECNOLOGÍA E INGENIERÍA

# Informe de Bases de datos

---

**Materia:** Proyecto Integrador  
**Carrera:** Tecnicatura en Inteligencia Artificial  
**Año:** 2025

---

**Alumno:** Nicolás Seivane  
**Profesora:** Andrea Rey

27 de agosto de 2025

# Informe de Data Sets Encontrados

Se realiza este informe de registros, atributos y métricas relevantes luego de eliminar duplicados y datos faltantes.

## 1º Desempeño de los Estudiantes

**Título Original:** Students Performance Dataset

**Link:** [Desempeño de los estudiantes](#)

**Tipo Licencia:** Attribution 4.0 International (CC BY 4.0)

**Tipo Archivo:** CV

**Descripción:** Este dataset contiene información comprensiva de 2392 estudiantes de secundaria, informando sobre datos demográfica, hábitos de estudio, el involucramiento parental, actividades extracurriculares y el desempeño académico. La variable objetivo que es, GradeClass, clasifica estudiantes en distintas categorías.

Es un dataset sintético

**Cantidad de registros:** 2392

**Cantidad de atributos:** 14

**Atributos Categóricos:** 10

**Atributos Numéricos:** 4

Los atributos son (Algunos son numéricos en el dataset pero son codificaciones de categóricos):

Atributo	Tipo de dato	¿Esta codificado?	Unidad
StudentID	Numérico (int)	No	-
Age	Numérico (int)	No	Años
Gender	Numérico (int)	Si	-
Ethnicity	Numérico (int)	Si	-
ParentalEducation	Numérico (int)	Si	-
StudyTimeWeekly	Numérico (float)	No	Horas
Absences	Numérico (int)	No	-
Tutoring	Numérico (int)	Si	-
ParentalSupport	Numérico (int)	Si	-
Extracurricular	Numérico (int)	Si	-
Sports	Numérico (int)	Si	-
Music	Numérico (int)	Si	-
Volunteering	Numérico (int)	Si	-
GPA	Numérico (float)	No	-
GradeClass	Entero (int)	Si	-

Descripción atributos:

**StudentID:** Identificador único asignado a cada alumno, rango de 1001 a 3392

**Age:** Edad en años de los estudiantes; hay 24.54 % de alumnos con 17, hay 24.33 % de alumnos con 18, hay 26.34 % de alumnos con 15 y hay 24.79 % de alumnos con 16.

**Gender:** Genero de los alumnos, donde hay 51.09 % de mujeres, codificadas en 1 y 48.91 % de hombres, codificados en 0.

**Ethnicity:** Etnia a la que pertenecen los alumnos; hay 50.46 % Caucásicos codificados con 0, 19.65 % Afroamericano codificados con 1, 20.61 % Asiáticos codificados con 2 y 9.28 % Otros codificados con 3.

**ParentalEducation:** Nivel Educativo de los padres; hay 10.16 % Sin estudios codificados con 0, hay 30.43 % Secundaria codificados con 1, hay 39.05 % Terciario codificados con 2, hay 15.34 % Grado codificados con 3, hay 5.02 % Posgrado codificados con 4.

**StudyTimeWeekly:** Es el tiempo semanal de estudio en horas, de rango 0 a 20, tiene media: 9.77, valor maximo: 19.98 y valor minimo: 0.00

**Absences:** Numero de Inasistencias durante el año escolar, rango de 0 a 30

**Tutoring:** Estado de tutoria; Hay 30.14 % Si codificado en 1 y hay 69.86 % No codificado en 0

**ParentalSupport:** El nivel de apoyo parental, hay 30.94 % Moderado codificado en 2, hay 20.44 % Bajo codificado en 1, hay 29.14 % Alto codificado en 3, hay 10.62 % Muy Alto codificado en 4, hay 8.86 % Nada codificado en 0

**Extracurricular:** Participación en Actividades Extracurriculares; hay 61.66 % No codificado en 0 y hay 38.34 % Si codificado en 1

**Sports:** Participación en Deportes; hay 69.65 % No codificado en 0 y hay 30.35 % Si codificado en 1

**Music:** Participacion en actividades Musicales; hay 80.31 % No codificado en 0 y hay 19.69 % Si codificado en 1

**Volunteering:** Participacion en voluntariados; hay 84.28 % No codificado en 0 y hay 15.72 % Si codificado en 1

**GPA:** Grade Point Average, califacion promedio en escala 2.0 a 4.0; tiene media: 1.91, valor maximo: 4.00 y valor minimo: 0.00

**GradeClass:** La variable objetivo, que posee; hay 4.47 % A codificados con 0, hay 11.25 % B codificados con 1, 16.35 % C codificados con 2, hay 17.31 % D con 3, hay 50.63 % F codificados con 4.

### Función Objetivo Inicial:

$$f(x) = \begin{cases} 'A' & \text{si } GPA \geq 3.5 \\ 'B' & \text{si } 3.0 \leq GPA < 3.5 \\ 'C' & \text{si } 2.5 \leq GPA < 3.0 \\ 'D' & \text{si } 2.0 \leq GPA < 2.5 \\ 'F' & \text{si } GPA < 2.0 \end{cases}$$

## 2º Insuficiencia Cardíaca Predicción

**Título Original:** Heart Failure Prediction Dataset

**Link:** [Insuficiencia Cardíaca](#)

**Tipo Licencia:** Database: Open Database, Contents: © Original Authors

**Tipo Archivo:** CV

**Descripción:** Enfermedades cardiovasculares son la causa numero uno de muerte globalmente, tomando un estimado de 17.9 millones de vidas cada año, que son aproximadamente 31 % de todas las muertes globales.(No se si poner algo asi)

Este dataset fue creado mediante la combinación de distintos dataset disponibles independientes pero no combinados anteriormente. 5 datasets de información cardíaca están combinados en 11 atributos comunes logrando el dataset mas grande de informacion de enfermedades cardiovasculares utilizado para investigación. Los 5 datasets utilizados para la creación de este son:

- Cleveland: 303 observaciones
- Hungarian: 294 observaciones
- Switzerland: 123 observaciones
- Long Beach VA: 200 observaciones
- Stalog (Heart) Data Set: 270 observaciones

**Cantidad de registros:** 918

**Cantidad de registros valiosos:** 743

**Cantidad de atributos:** 12

**Atributos Categóricos:** 5

**Atributos Numéricos:** 7

Los atributos son (Algunos son numéricos en el dataset pero son codificaciones de categóricos):

Atributo	Tipo de dato	¿Esta codificado?	Unidad
Age	Numérico (int)	No	Años
Sex	Categorico (string)	No	-
ChestPainType	Categorico (string)	No	-
RestingBP	Numérico (int)	No	mm Hg
Cholesterol	Numérico (int)	No	mm/dl
FastingBS	Numérico (int)	Si	"mg/dl"
RestingECG	Categorico (string)	No	-
MaxHR	Numérico (int)	No	-
ExerciseAngina	Categorico (string)	No	-
Oldpeak	Numérico (float)	No	ST en depresión
ST_Slope	Categorico (string)	No	-
HeartDisease	Numérico (int)	Si	-

Descripción atributos:

**Age:** Edad de los pacientes con media: 53 años, valor máximo: 77 y valor mínimo: 28, con proporciones de edad bastante bien distribuidas, siendo la menor de 0.11 % para algunas edades y la mayor de 4.14 % para otras edades, teniendo otras distribuciones entre estos dos rangos

**Sex:** Sexo de los pacientes; hay 78.98 % M (masculinos) y hay 21.02 % F (femeninos)

**ChestPainType:** Tipo del dolor en el pecho; Hay 18.85 % ATA, hay 22.11 % NAP, hay 54.03 % ASY, hay 5.01 % TA. [TA: Typical Angina, ATA: Atypical Angina, NAP: Non-Anginal Pain, ASY: Asymptomatic]

**RestingBP:** Presión sanguínea en reposo, donde hay 51.09 % de mujeres, codificadas en 1 y 48.91 % de hombres, codificados en 0.

**Cholesterol:** Colesterol serico, la medida total de colesterol en sangre; tiene media: 199.02, valor máximo: 603.00 y valor mínimo: 0.00. Miligramos por decilitro

**FastingBS:** Glucosa en sangre en ayuno; hay 76.66 % Glucosa en sangre  $\geq 120$  mg/dl codificado en 0 y hay 23.34 % Glucosa en sangre  $\geq 120$  mg/dl codificado en 1

**RestingECG:** Resultados de electrocardiogramas en reposo; hay 60.09 % codificado en Normal, hay 19.41 % codificado en ST y hay 20.50 % codificado en LVH [Normal: Normal, ST: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of  $\geq 0.05$  mV), LVH: showing probable or definite left ventricular hypertrophy by Estes' criteria]

**MaxHR:** Máximo ritmo cardíaco registrado, tiene media: 136.79, valor maximo: 202.00 y valor minimo: 60.00

**ExerciseAngina:** Angina producido por ejercicio, dolor en el pecho; hay 59.54 % No codificado en N y hay 40.46 % Si codificado en Y

**Oldpeak:** Valor máximo de depresión del segmento ST (en milímetros) registrado en todas las derivaciones contiguas durante una prueba de esfuerzo. Forma parte del cálculo del riesgo de un paciente de isquemia o infarto de miocardio; valores más altos indican un mayor riesgo de enfermedad coronaria; tiene media: 0.90, valor maximo: 6.20 y valor minimo: -0.10

**ST\_Slope:** The slope of the peak exercise ST segment; hay 43.08 %, hay 50.05 % Flat y hay 6.87 % Down [Up: upsloping, Flat: flat, Down: downsloping]

**HeartDisease:** Variable de salida de si posee una enfermedad cardíaca; hay 44.71 % No codificado en 0 y hay 55.29 % Si codificado en 1

**Función Objetivo Inicial:** Donde la variable salida es *HeartDisease*, no hay una variable que se use como condición:

$$f(x) = \begin{cases} '1' & \text{si } ?? \\ '0' & \text{si } ?? \end{cases}$$

### 3º Predicción de Diabetes

**Título Original:** Diabetes Prediction Dataset

**Link:** [Predicción de Diabetes](#)

**Tipo Licencia:** Attribution 4.0 International (CC BY 4.0)

**Tipo Archivo:** CV

**Descripción:** Este dataset parece estar relacionado a la diabetes y contiene varias medidas medicas y características de los pacientes.

**Cantidad de registros:** 1000

**Cantidad de atributos:** 14

**Atributos Categóricos:** 2

**Atributos Numéricos:** 12

Los atributos son :

Atributo	Tipo de dato	¿Esta codificado?	Unidad
ID	Numérico (int)	No	-
<i>No_Pation</i>	Numérico (int)	No	-
Gender	Categorico (str)	No	-
AGE	Numérico (int)	No	Años
Urea	Numérico (float)	No	mg/dL
Cr	Numérico (float)	No	mg/dL
HbA1c	Numérico (float)	No	-
Chol	Numérico (float)	No	mg/dL
TG	Numérico (float)	Si	-
HDL	Numérico (float)	No	mg/dL
LDL	Numérico (float)	No	mg/dL
VLDL	Numérico (float)	No	mg/dL
BMI	Numérico (float)	No	-
CLASS	Categorico (str)	Si	-

Descripción atributos:

**ID:** Identificador de cada registro, de 0 a 999.

*No\_Pation*: Otra forma de identificación del paciente

**Gender:** Genero de los pacientes, donde hay 43.50 % de mujeres, codificadas en 1 y 56.60 % de hombres, codificados en 0.

**AGE:** Edad de los pacientes: tiene media: 53 años, valor máximo: 79 y valor mínimo: 20

**Urea:** Nivel de urea en sangre, es un producto de desecho del metabolismo de las proteínas y puede indicar la función renal; tiene media: 5.12, valor máximo: 38.90 y valor mínimo: 0.50

**Cr:** Nivel de creatinina en sangre (probablemente medido en mg/dL o  $\mu\text{mol/L}$ ). La creatinina es otro producto de desecho que indica la función renal; tiene media: 68.94, valor máximo: 800.00 y valor mínimo: 6.00

**HbA1c:** Hemoglobina glucosilada, una medida de los niveles promedio de azúcar en sangre durante los últimos 2-3 meses (expresada como porcentaje); tiene media: 8.28, valor maximo: 16.00 y valor minimo: 0.90

**Chol:** Nivel de colesterol en sangre (probablemente medido en mg/dL o mmol/L). Esto generalmente se refiere al colesterol total; tiene media: 4.86, valor maximo: 10.30 y valor minimo: 0.00

**TG:** Nivel de triglicéridos en sangre (probablemente medido en mg/dL o mmol/L). Los triglicéridos son un tipo de grasa presente en la sangre; tiene media: 2.35, valor maximo: 13.80 y valor minimo: 0.30

**HDL:** Nivel de colesterol de lipoproteínas de alta densidad (a menudo llamado colesterol "bueno", medido en mg/dL o mmol/L); tiene media: 1.20, valor maximo: 9.90 y valor minimo: 0.20

**LDL:** Nivel de colesterol de lipoproteínas de baja densidad (a menudo llamado colesterol "malo", medido en mg/dL o mmol/L); tiene media: 2.61, valor maximo: 9.90 y valor minimo: 0.30

**VLDL:** Nivel de colesterol de lipoproteínas de muy baja densidad (medido en mg/dL o mmol/L); tiene media: 1.85, valor maximo: 35.00 y valor minimo: 0.10

**BMI:** Índice de Masa Corporal, una medida de la grasa corporal basada en la altura y el peso (calculado como el peso en kilogramos dividido por la altura en metros al cuadrado); tiene media: 29.58, valor maximo: 47.75 y valor minimo: 19.00

**CLASS:** La etiqueta de clase que indica el estado de diabetes del paciente. Los valores posibles parecen ser; N: No diabético, P: Prediabético y Y: Diabético. Hay 10.30 % N, hay 5.30 % P y hay 84.40 % Y

**Función Objetivo Inicial:** El atributo objetivo es *CLASS*

$$f(x) = \begin{cases} \text{'N'} & \text{si ?} \\ \text{'P'} & \text{si ?} \\ \text{'Y'} & \text{si ?} \end{cases}$$

## 4º Evaluación de la calidad del aire y la contaminación

**Título Original:** Air Quality and Pollution Assessment

**Link:** Evaluación de la calidad del aire y la contaminación

**Tipo Licencia:** Apache 2.0

**Tipo Archivo:** CV

**Descripción:** Contiene evaluación de la calidad del aire a través de varias regiones. Es un dataset que contiene 5000 muestras y captura factores críticos ambientales y demográficos que influyen los niveles de contaminación. Las fuentes son; World Health Organization (WHO) (<https://www.who.int/health-topics/air-pollution>) y World Bank Data (<https://data.worldbank.org/indicator/EN.POP.DNST>)

**Cantidad de registros:** 5000

**Cantidad de atributos:** 10

**Atributos Categóricos:** 1

**Atributos Numéricos:** 9

Los atributos son :

Atributo	Tipo de dato	¿Esta codificado?	Unidad
Temperature	Numérico (float)	No	°C
Humidity	Numérico (float)	No	-
PM2.5	Numérico (float)	No	g/m
PM10	Numérico (float)	No	g/m
NO2	Numérico (float)	No	ppb
SO2	Numérico (float)	No	ppb
CO	Numérico (float)	No	ppm
<i>Proximity_to_Industrial_Areas</i>	Numérico (float)	No	km
<i>Population_Density</i>	Numérico (int)	No	personas/km <sup>2</sup>
Air Quality	Categórico (str)	No	-

Descripción atributos:

**Temperature:** Temperatura promedio de la región; tiene media: 30.03, valor maximo: 58.60 y valor minimo: 13.40

**Humidity:** Humedad relativa registrada en la región; tiene media: 70.06, valor maximo: 128.10 y valor minimo: 36.00

**PM2.5:** Niveles de partículas finas; tiene media: 20.14, valor maximo: 295.00 y valor minimo: 0.00

**PM10:** Niveles de partículas gruesas; tiene media: 30.22, valor maximo: 315.80 y valor minimo: -0.20

**NO2:** Niveles de dióxido de nitrógeno; tiene media: 26.41, valor maximo: 64.90 y valor minimo: 7.40

**SO2:** Niveles de dióxido de azufre; tiene media: 10.01, valor maximo: 44.90 y valor minimo: -6.20

**CO:** Niveles de monóxido de carbono; tiene media: 1.50, valor maximo: 3.72 y valor minimo: 0.65

**Proximity\_to\_Industrial\_Areas:** Distancia a la zona industrial más cercana; tiene media: 8.43, valor maximo: 25.80 y valor minimo: 2.50

**Population\_Density:** Número de personas por kilómetro cuadrado en la región; tiene media: 497, valor maximo: 957.00 y valor minimo: 188.00

**Air Quality :** Niveles de Calidad del Aire

Buena: Aire limpio con bajos niveles de contaminación.

Moderada: Calidad del aire aceptable, pero con presencia de algunos contaminantes.

Deficiente: Contaminación considerable que puede causar problemas de salud a grupos vulnerables.



Peligroso: Aire altamente contaminado que supone graves riesgos para la salud de la población.

Hay 30.00 % Moderado, hay 40.00 % Buena, hay 10.00 % Peligroso y hay 20.00 % Deficiente

**Función Objetivo Inicial:** El atributo objetivo es *AirQuality*

$$f(x) = \begin{cases} \text{'Buena'} & \text{si ?} \\ \text{'Moderada'} & \text{si ?} \\ \text{'Deficiente'} & \text{si ?} \\ \text{'Peligroso'} & \text{si ?} \end{cases}$$

## 5º Desempeño corporal

**Título Original:** Body performance Data

**Link:** Desempeño corporal

**Tipo Licencia:** CC0: Public Domain

**Tipo Archivo:** CV

**Descripción:** Esta es la data que confirma el grado de desempeño con la edad y algunos ejercicios realizados.

Link coreano original(Korea Sports Promotion Foundation) El dataset actual esta procesado del original..

**Cantidad de registros:** 13392

**Cantidad de atributos:** 12

**Atributos Categóricos:** 2

**Atributos Numéricos:** 10

Los atributos son :

Atributo	Tipo de dato	¿Esta codificado?	Unidad
age	Numérico (float)	No	Años
gender	Categórico (str)	No	-
height_cm	Numérico (float)	No	cm
weight_kg	Numérico (float)	No	kg
bodyfat_%	Numérico (float)	No	%
diastolic	Numérico (float)	No	min
systolic	Numérico (float)	No	min
gripForce	Numérico (float)	No	kg
sitandbendforward_cm	Numérico (float)	No	cm
sit-ups counts	Numérico (float)	No	Cantidad
broadjump_cm	Numérico (float)	No	cm
class	Categórico (str)	No	-

Descripción atributos:

**age:** Edad de los participantes; tiene media: 36, valor maximo: 64 y valor minimo: 21

**gender:** Sexo de los participantes; hay 63.22 % Hombres codificado en M y hay 36.78 % Mujeres codificado en F

**height\_cm:** Altura de los participantes; tiene media: 168.56, valor maximo: 193.80 y valor minimo: 125.00

**weight\_kg:** Peso de los participantes; tiene media: 67.45, valor maximo: 138.10 y valor minimo: 26.30

**body fat %:** Niveles de grasa corporal de los participantes; tiene media: 23.24, valor maximo: 78.40 y valor minimo: 3.00

**diastolic:** Presión arterial diastólica; tiene media: 78.80, valor maximo: 156.20 y valor minimo: 0.00

**systolic:** Presión arterial sistólica; tiene media: 130.24, valor maximo: 201.00 y valor minimo: 0.00

**gripForce:** Fuerza de agarre; tiene media: 36.96, valor maximo: 70.50 y valor minimo: 0.00

**sit and bend forward\_cm:** Sentarse e inclinarse hacia adelante; tiene media: 15.21, valor maximo: 213.00 y valor minimo: -25.00

**sit-ups counts:** Recuento de abdominales; tiene media: 39.77, valor maximo: 80.00 y valor minimo: 0.00

**broad jump\_cm:** Salto en Alto; tiene media: 190.13, valor maximo: 303.00 y valor minimo: 0.00

**class:** Clase; Hay 25.01 % C, hay 24.99 % A, hay 24.99 % B y hay 25.01 % D

**Función Objetivo Inicial:** El atributo objetivo es *class*

$$f(x) = \begin{cases} 'A' & \text{si ?} \\ 'B' & \text{si ?} \\ 'C' & \text{si ?} \\ 'D' & \text{si ?} \end{cases}$$

## 5º Microbios

**Título Original:** Bank Marketing

**Link:** [Microbios](#)

**Tipo Licencia:** CC0: Public Domain

**Tipo Archivo:** CV

**Descripción:** Las nuevas tecnologías de secuenciación de ADN han proliferado en las últimas dos décadas. Las mejoras continuas en la secuenciación de nueva generación (NGS) y la secuenciación de tercera generación (TGS) han aumentado la fidelidad y la velocidad de la secuenciación, pero aún se necesitan horas o días para obtener secuencias completas. Además, existen algunas aplicaciones diagnósticas en las que la identificación rápida de un gen o especie genética en particular se vuelve esencial, mientras que la identificación de

todos los genes no es necesaria. Por ejemplo, en pacientes con shock séptico por infecciones bacterianas, la identificación de genes de resistencia a los antibióticos es esencial porque la tasa de mortalidad aumenta un 7,6 % por cada hora de retraso en la administración de los antibióticos correctos. Desafortunadamente, se necesitan más de 24 horas para cultivar las bacterias recuperadas de la sangre de un paciente infectado, identificar la especie y luego determinar a qué antibióticos es resistente el organismo, lo que resulta en una tasa de mortalidad muy alta en estas infecciones.

La resistencia bacteriana a los antibióticos se está convirtiendo en una amenaza importante para la salud, y la rápida identificación de bacterias resistentes a los antibióticos es esencial para salvar vidas y reducir su propagación.

**Cantidad de registros:** 30527

**Cantidad de atributos:** 26

**Atributos Categóricos:** 1

**Atributos Numéricos:** 25

Los atributos son :

Atributo	Tipo de dato	¿Esta codificado?	Unidad
Unnamed	Numérico (int)	No	-
Solidity	Numérico (float)	No	-
Eccentricity	Numérico (float)	No	-
EquivDiameter	Numérico (float)	No	-
Extrema	Numérico (float)	No	-
FilledArea	Numérico (float)	No	-
Extent	Numérico (float)	No	-
Orientation	Numérico (float)	No	Grados
EulerNumber	Numérico (float)	No	-
BoundingBox1	Numérico (float)	No	-
BoundingBox2	Numérico (float)	No	-
BoundingBox3	Numérico (float)	No	-
BoundingBox4	Numérico (float)	No	-
ConvexHull1	Numérico (float)	No	-
ConvexHull2	Numérico (float)	No	-
ConvexHull3	Numérico (float)	No	-
ConvexHull4	Numérico (float)	No	-
MajorAxisLength	Numérico (float)	No	-
MinorAxisLength	Numérico (float)	No	-
Perimeter	Numérico (float)	No	-
ConvexArea	Numérico (float)	No	-
Centroid1	Numérico (float)	No	-
Centroid2	Numérico (float)	No	-
Area	Numérico (float)	No	-
raddi	Numérico (float)	No	-
microorganisms	Categórico (str)	No	-

Descripción atributos:

**Unnamed:** Índice de los registro

**Solidity:** Es la razón entre el área de un objeto y el área de su envoltura convexa. Se calcula como  $\text{Área}/\text{ÁreaConvexa}$ ; tiene media: 9.68, valor maximo: 23.00 y valor minimo: 0.00

**Eccentricity:** La excentricidad es la razón entre la longitud del eje mayor y el eje menor de un objeto; tiene media: 19.47, valor maximo: 23.00 y valor minimo: 0.00

**EquivDiameter:** Diámetro de un círculo con la misma área que la región; tiene media: 3.63, valor maximo: 23.00 y valor minimo: 0.00

**Extrema:** Puntos extremos en la región. El formato del vector es [arriba-izquierda arriba-derecha derecha-arriba derecha-abajo abajo-derecha abajo-izquierda izquierda-abajo izquierda-arriba]; tiene media: 11.87, valor maximo: 23.00 y valor minimo: 0.00

**FilledArea :** Número de píxeles en la ImagenRellenada, devuelto como un escalar; tiene media: 0.42, valor maximo: 23.00 y valor minimo: 0.00

**Extent:** Razón entre el área de píxeles de una región y el área del cuadro delimitador de un objeto; tiene media: 5.84, valor maximo: 23.00 y valor minimo: 0.00

**Orientation:** Dirección general de la forma. El valor varía de -90 a 90 grados; tiene media: 11.75, valor maximo: 23.00 y valor minimo: 0.00

**EulerNumber:** Número de objetos en la región menos el número de agujeros en esos objetos; tiene media: 22.38, valor maximo: 23.00 y valor minimo: 0.00

**BoundingBox1:** Posición y tamaño del rectángulo más pequeño que delimita el objeto; tiene media: 10.92, valor maximo: 23.00 y valor minimo: 0.00

**BoundingBox2:** Cuadro Delimitador 2; tiene media: 10.40, valor maximo: 23.00 y valor minimo: 0.00

**BoundingBox3:** Cuadro Delimitador 3; tiene media: 2.09, valor maximo: 23.00 y valor minimo: 0.00

**BoundingBox4:** Cuadro Delimitador 4; tiene media: 2.64, valor maximo: 23.00 y valor minimo: 0.00

**ConvexHull1:** Cubierta Convexa 1; tiene media: 11.11, valor maximo: 23.00 y valor minimo: 0.00

**ConvexHull2:** Cubierta Convexa 2; tiene media: 11.11, valor maximo: 23.00 y valor minimo: 0.00

**ConvexHull3:** Cubierta Convexa 3; tiene media: 11.05, valor maximo: 23.00 y valor minimo: 0.00

**ConvexHull4:** Cubierta Convexa 4; tiene media: 11.02, valor maximo: 23.00 y valor minimo: 0.00

**MajorAxisLength:** Longitud del Eje Mayor; tiene media: 1.61, valor maximo: 23.00 y valor minimo: 0.00

**MinorAxisLength:** Longitud del Eje Menor; tiene media: 1.01, valor maximo: 23.00 y valor minimo: 0.00

**Perimeter:** Perímetro; tiene media: 0.83, valor maximo: 23.00 y valor minimo: 0.00

**ConvexArea:** Área Convexa; tiene media: 0.25, valor maximo: 23.00 y valor minimo: 0.00

**Centroid1:** Centroide 1; tiene media: 11.75, valor maximo: 23.00 y valor minimo: 0.00

**Centroid2:** Centroide 2; tiene media: 11.55, valor maximo: 23.00 y valor minimo: 0.00

**Area:** Área; tiene media: 0.80, valor maximo: 23.00 y valor minimo: 0.00

**raddi:** Raddi; tiene media: 5.21, valor maximo: 23.00 y valor minimo: 0.00

**microorganisms:** 2.00 % Spirogyra  
 14.15 % Volvox  
 4.42 % Pithophora  
 11.79 % Yeast  
 8.36 % Raizopus  
 3.54 % Penicillum  
 12.74 % Aspergillus sp  
 12.74 % Protozoa  
 5.96 % Diatom  
 24.31 % Ulothrix

**Función Objetivo Inicial:** El atributo objetivo es *microorganisms*

$$f(x) = \begin{cases} \text{'microorganisms'} & \text{si ?} \\ \text{'Spirogyra'} & \text{si ?} \\ \text{'Volvox'} & \text{si ?} \\ \text{'Pithophora'} & \text{si ?} \\ \text{'Yeast'} & \text{si ?} \\ \text{'Raizopus'} & \text{si ?} \\ \text{'Penicillum'} & \text{si ?} \\ \text{'Aspergillus sp'} & \text{si ?} \\ \text{'Protozoa'} & \text{si ?} \\ \text{'Diatom'} & \text{si ?} \\ \text{'Ulothrix'} & \text{si ?} \end{cases}$$

Drive con Links