

Aplicación Interactiva

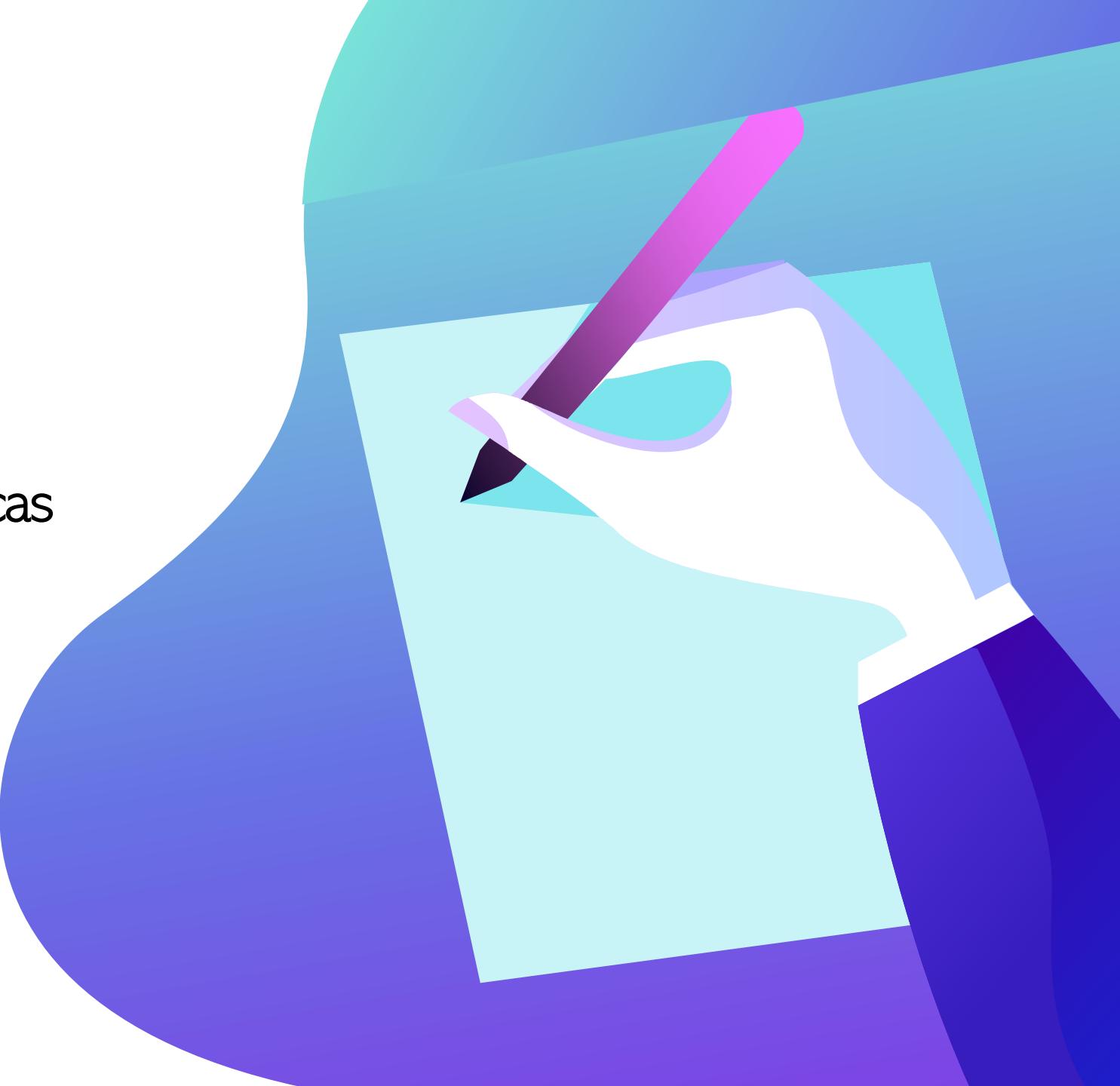
**para el análisis de
datos mediante**

**mapas
autoorganizados.**

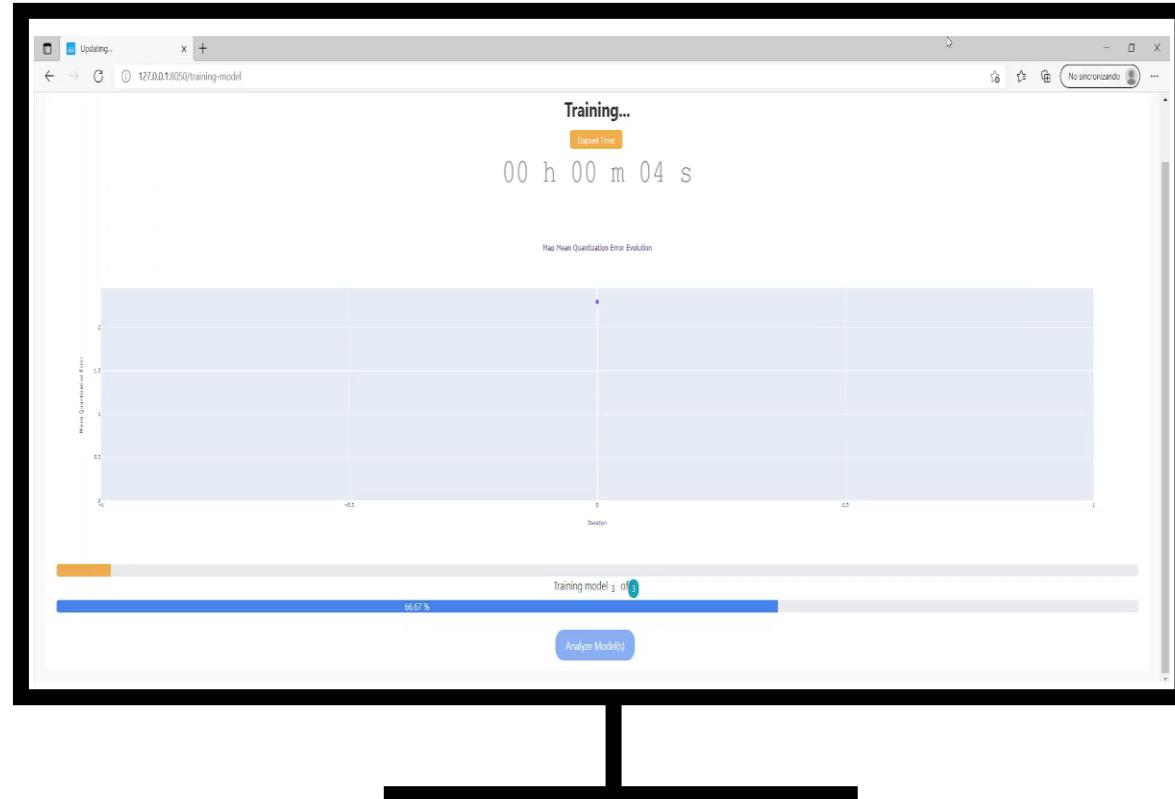
https://github.com/adriwitek/som_analytic_tool

ÍNDICE

1. Introducción
2. Estado del Arte
3. Diseño
4. Funcionalidades Analíticas
5. Implementaciones
6. Casos de Estudio
7. Trabajo Futuro
8. Conclusiones
9. Referencias de Figuras



Aplicación Web



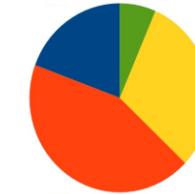
Diseño Responsive



Totalmente Visual

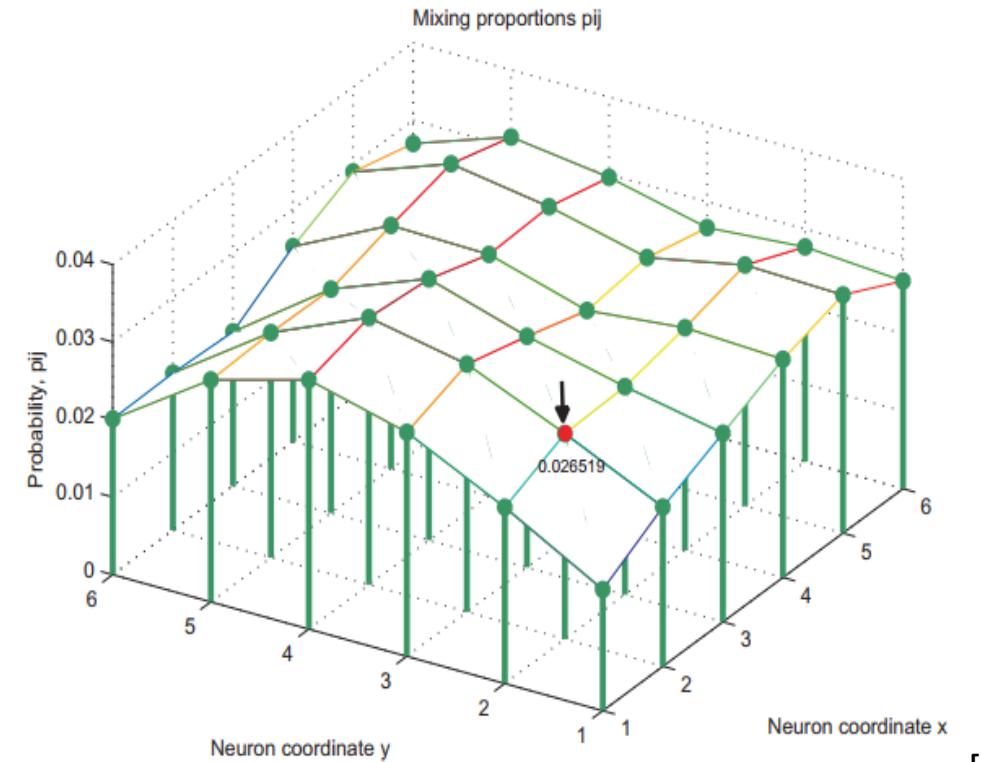
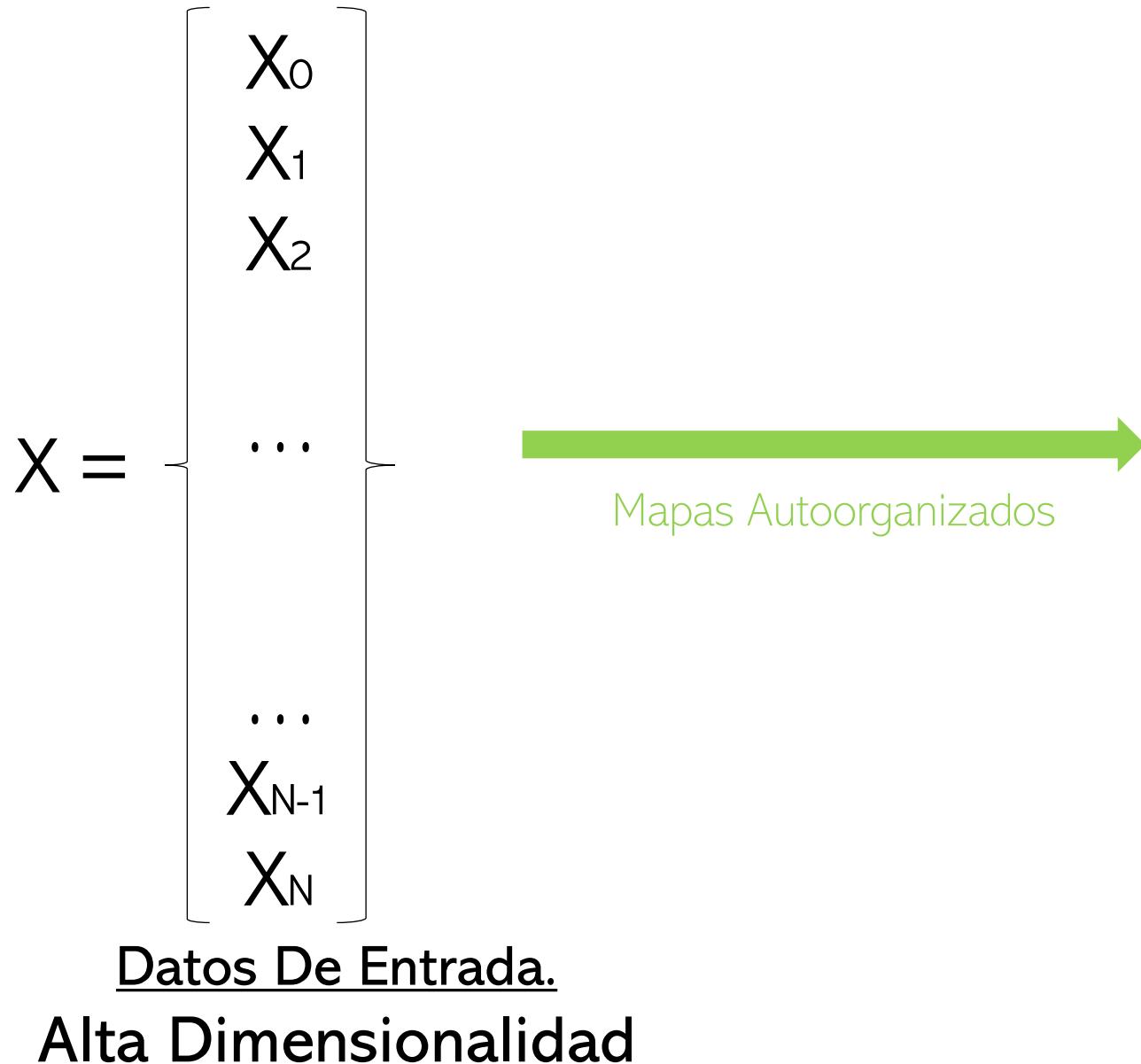


Interactiva



Analítica de Datos

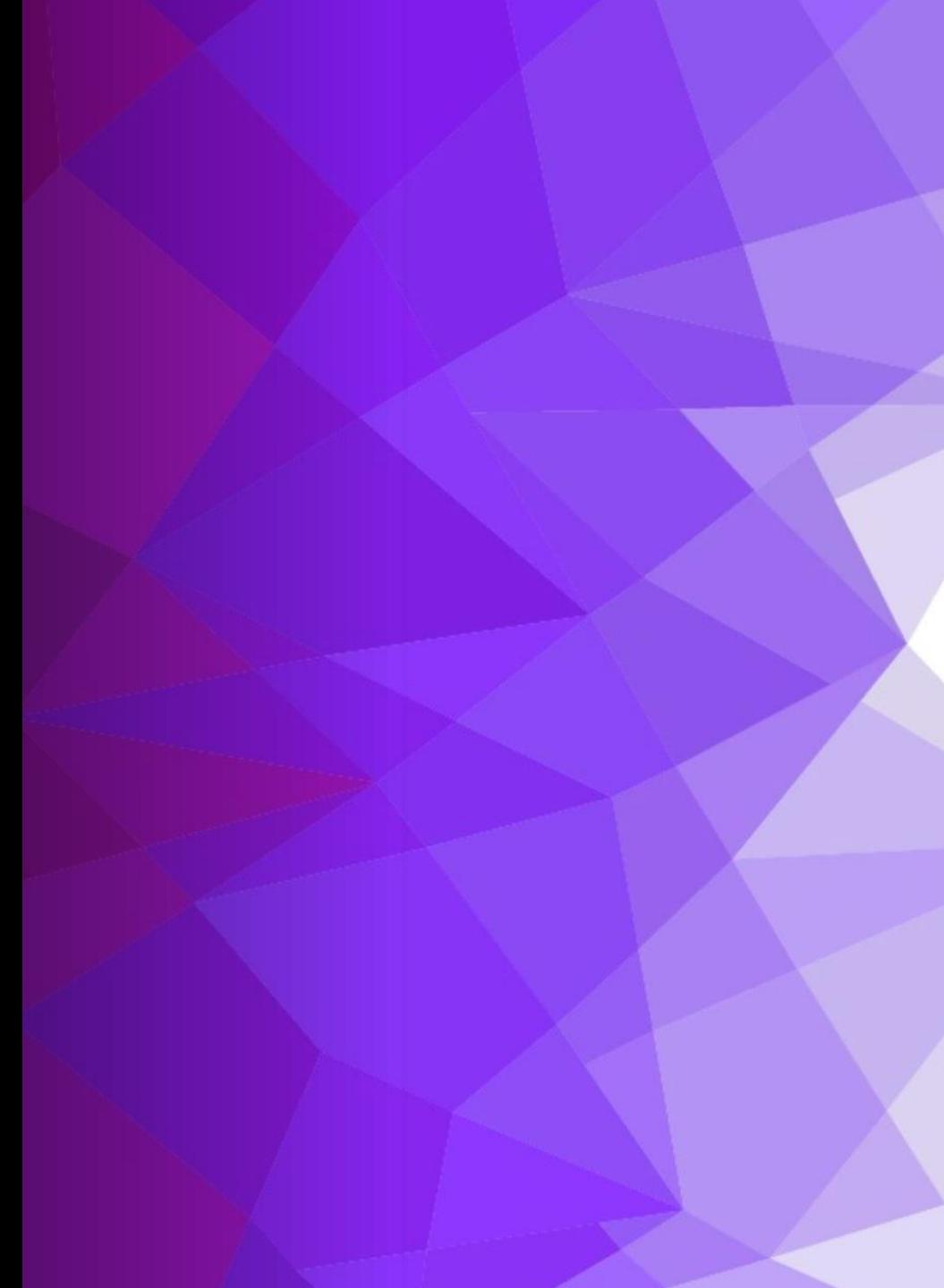
Reducción de la Dimensionalidad



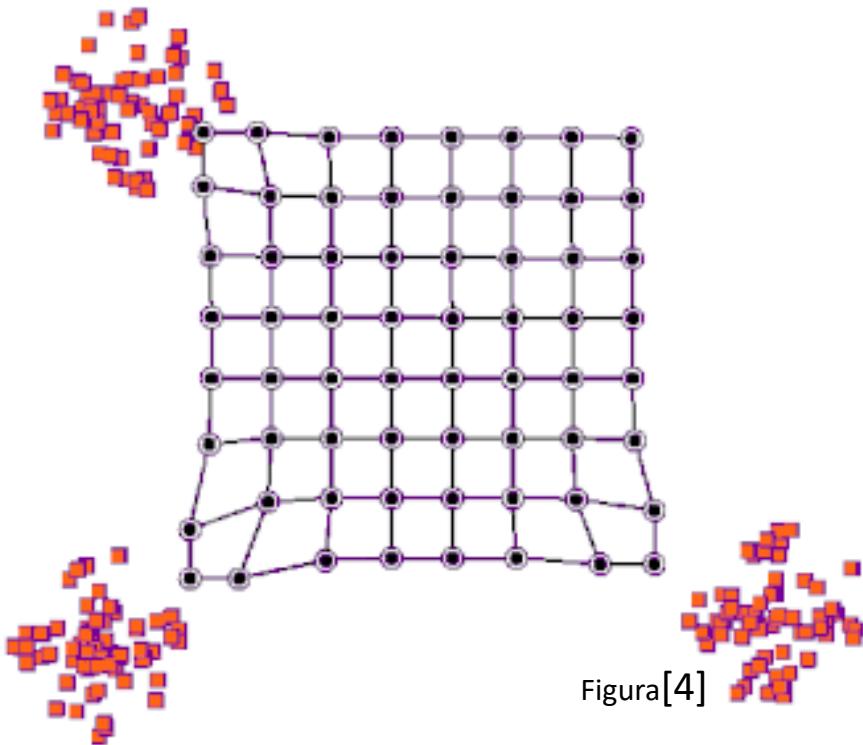
Datos Salida.
2 Dimensiones



ESTADO DEL ARTE

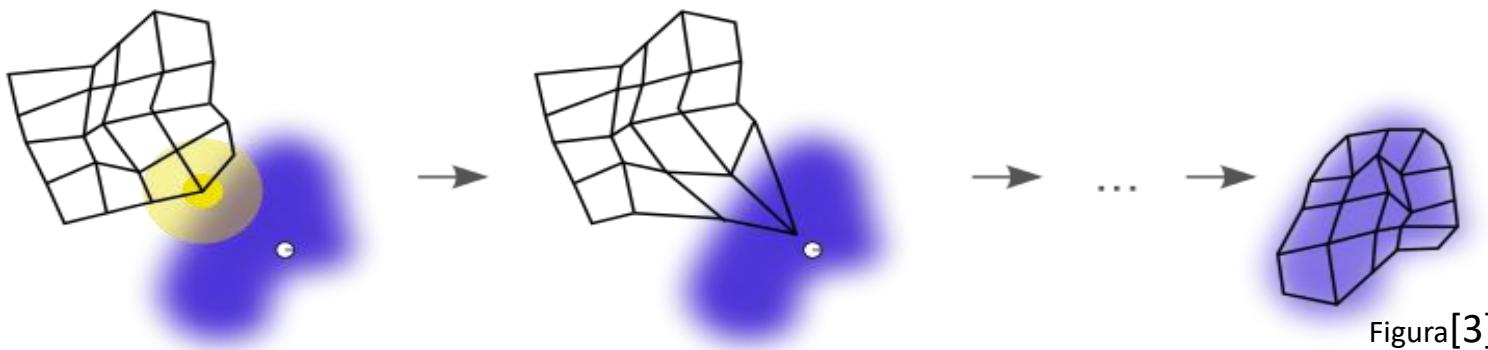


SOM's: Mapas Autoorganizados

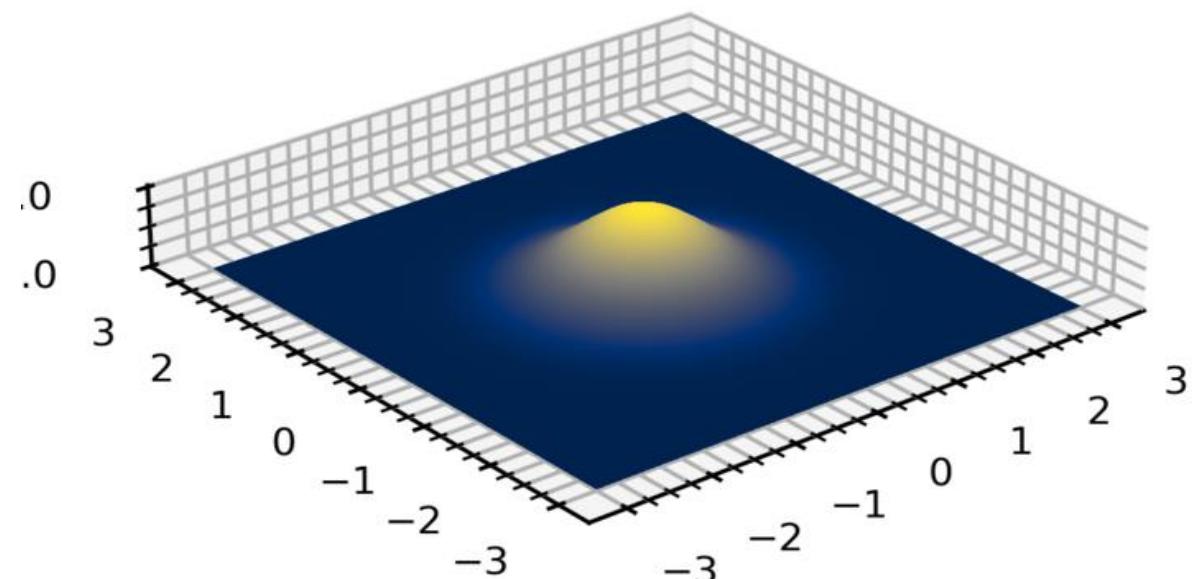
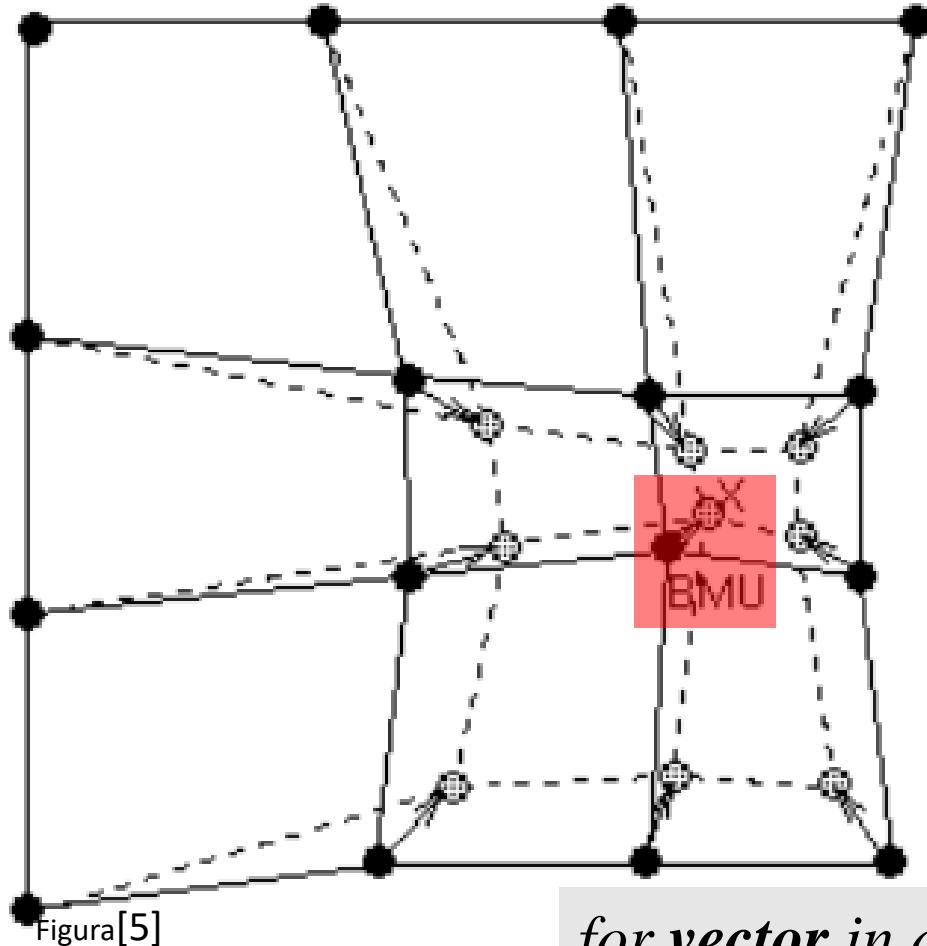


- Aprendizaje No Supervisado y Competitivo
- Reducción de la Dimensionalidad

- Detección de patrones y anomalías
- Clustering



SOM's: Entrenamiento



Figura[6]

for vector in datos_entrenamiento:

1. **BMU** = calcular_bmu(vector, mapa)
2. actualizar_pesos_mapa(BMU, vecindario_BMU)

SOM's: Entrenamiento

Actualización de los pesos de una neurona.

$$w_i(t + 1) = \underline{w_i(t)} + \underline{\alpha(t)} \cdot \underline{h_{ci}(t)} \cdot \underline{x(t)} - \underline{w_i(t)}$$

Nuevo
Vector de
Pesos

Vector de
Pesos Actual



Tasa de
Aprendizaje



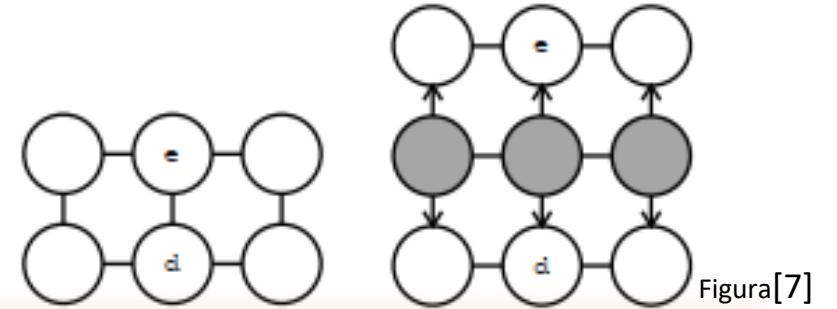
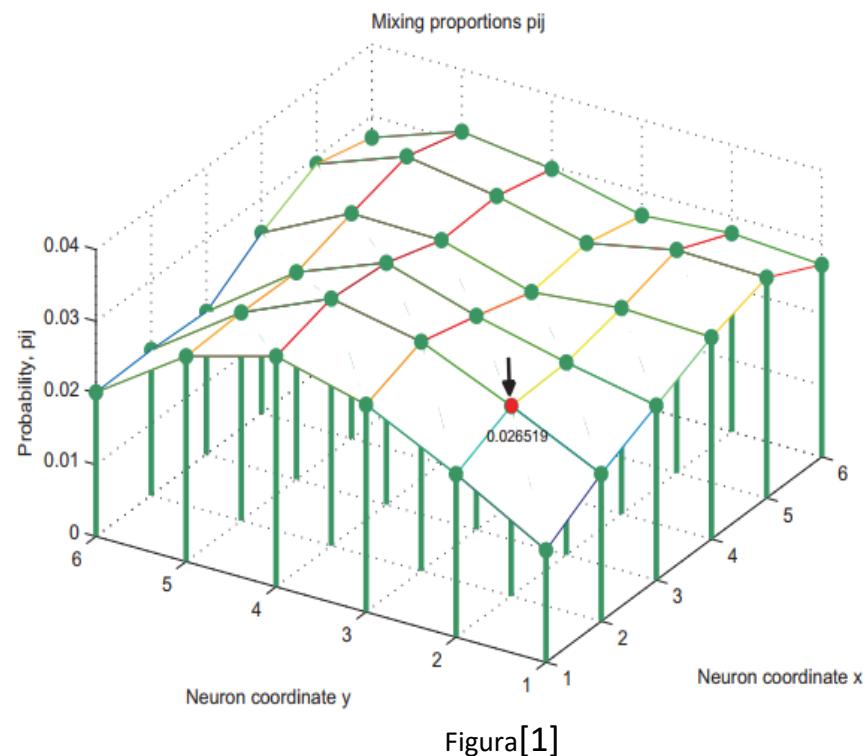
Vector
Entrada
Pesos Actual

Función de Vecindad

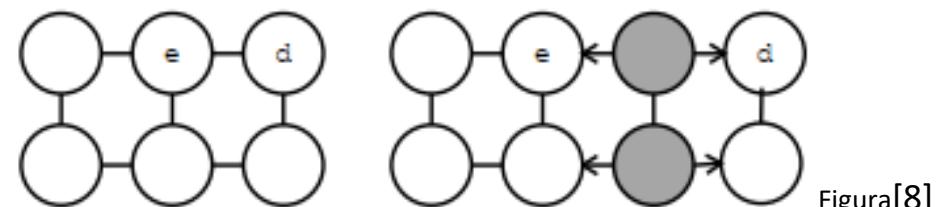
for vector in datos_entrenamiento:

1. **BMU** = calcular_bmu(vector, mapa)
2. actualizar_pesos_mapa(BMU, vecindario_BMU)

GSOM: SOM's Crecientes



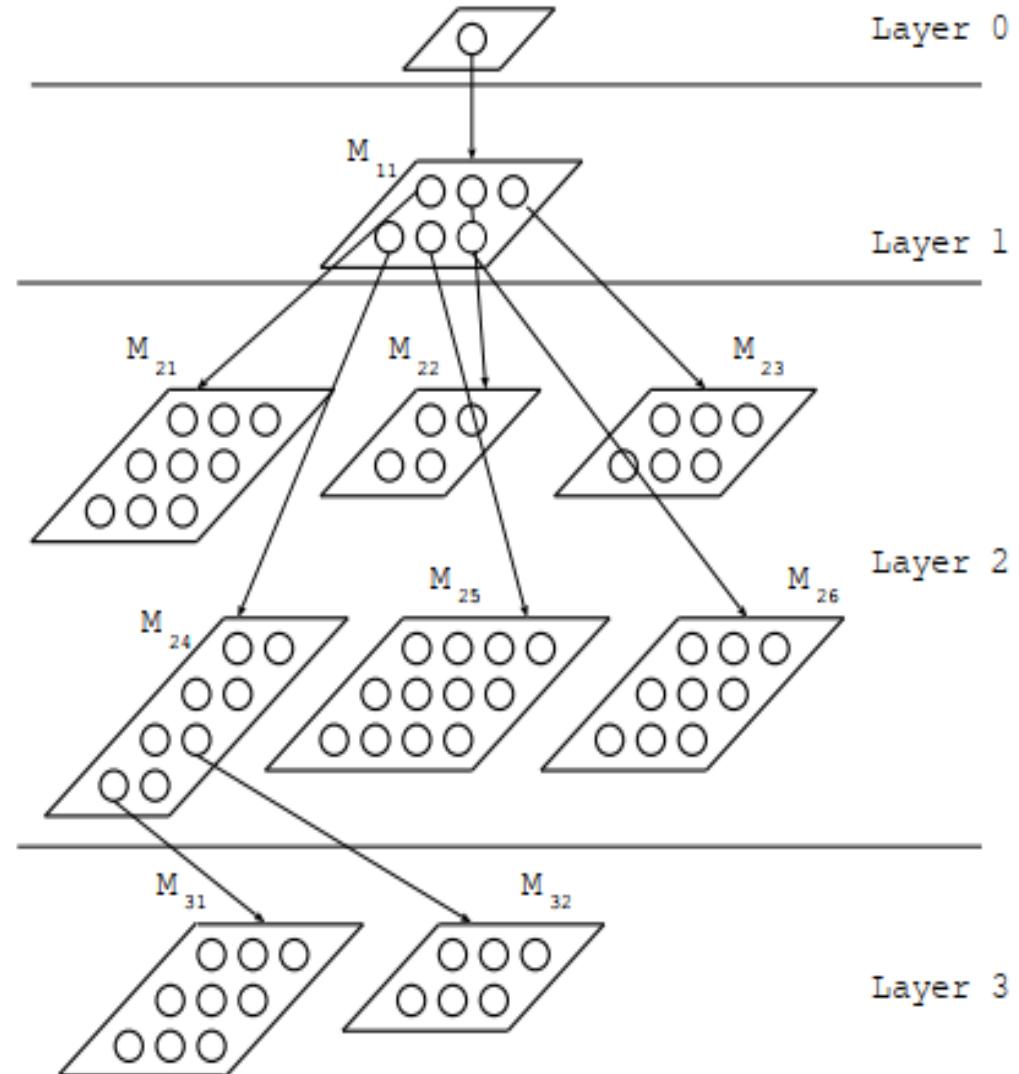
Inserción de una Fila



Inserción de una
Columna

GHSOM's: GSOM's Jeráquicos

Orientación
Global de la Red



Figura[9]



DISEÑO

Selección de librerías

Plataformas

Python 3.8

Dash

...



Algoritmos Ya Implementados

“Minisom”



SOM

Algoritmos

“The Growing Hiearchical Self-Organizing Map”



GSOM



GHSOM

Desarrollos y ajustes propios

Desarrollo Propio

Plataformas

Python 3.8

Dash

...

Web App

Diagrama de Flujo de uso de la Aplicación

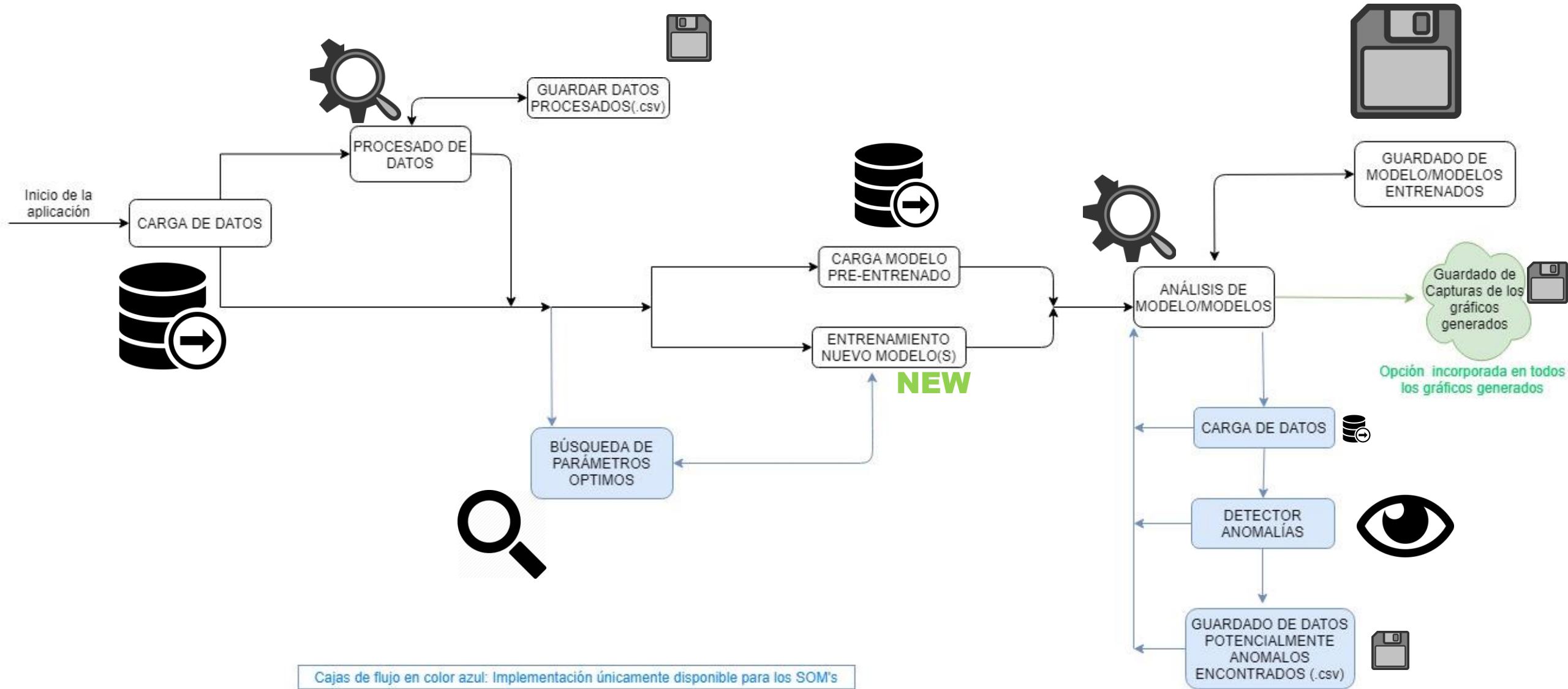
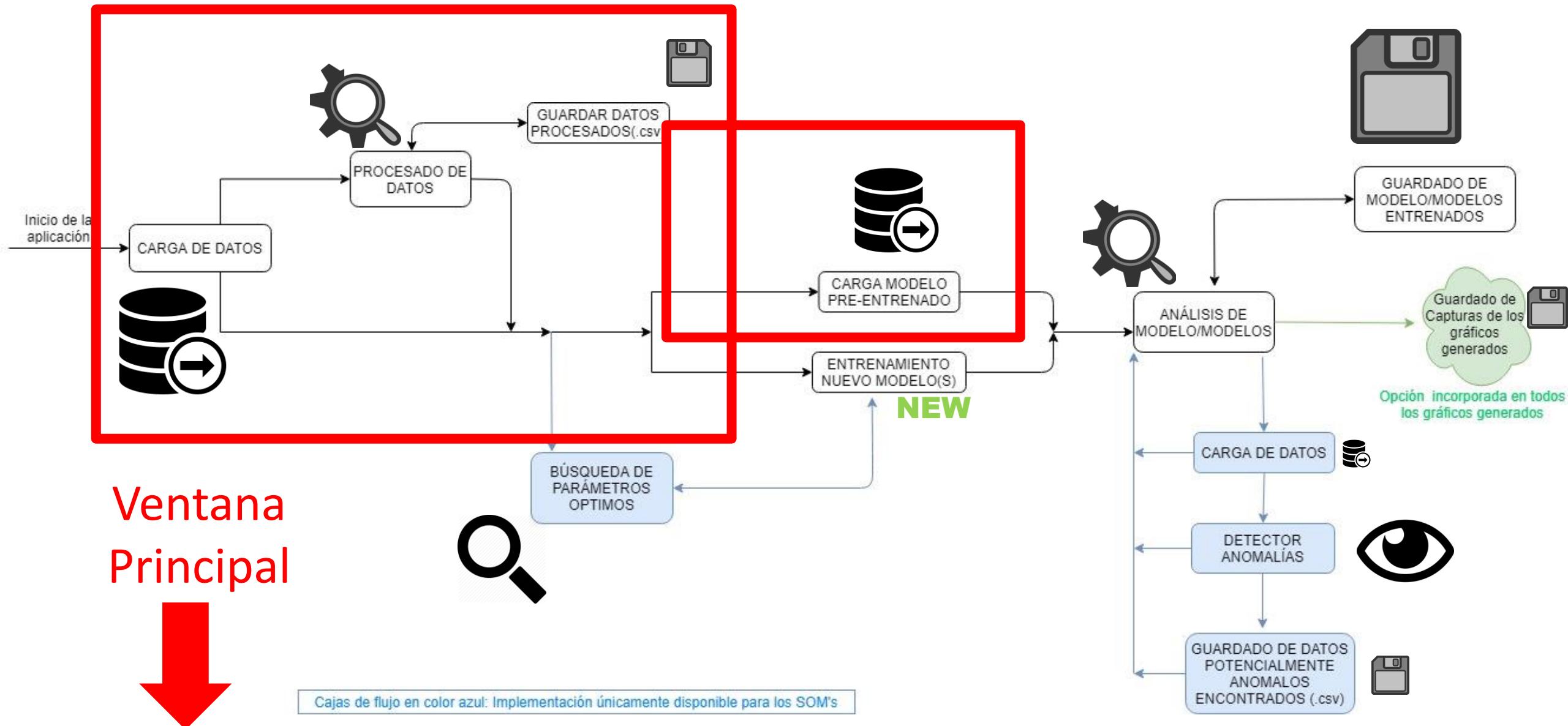


Diagrama de Flujo de uso de la Aplicación



VENTANA DE BIENVENIDA

SOM Analytic Tool

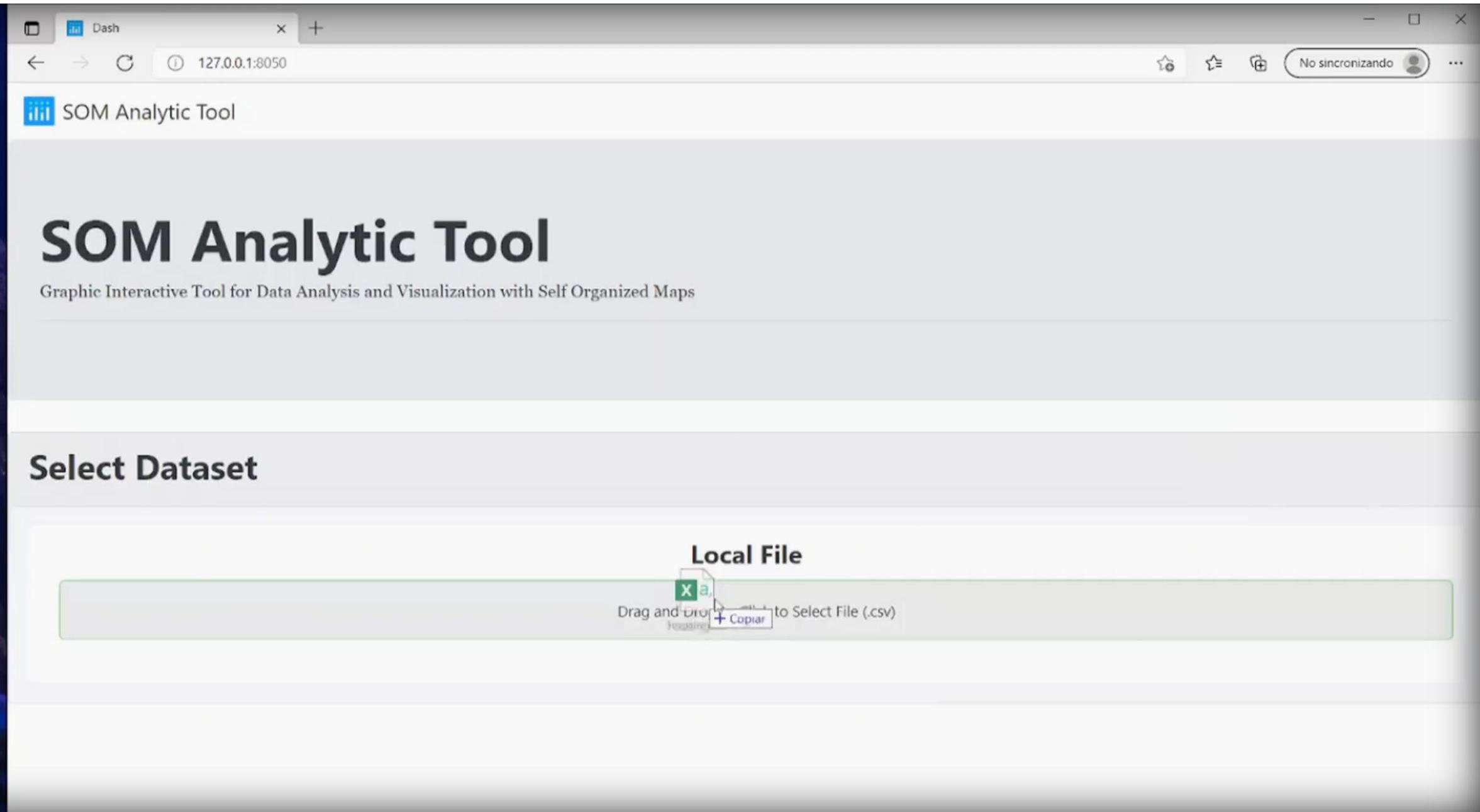
SOM Analytic Tool

Graphic Interactive Tool for Data Analysis and Visualization with Self Organized Maps

Select Dataset

Local File

Drag and drop  to Select File (.csv)
+ Copiar



FUNCIONALIDADES ANALÍTICAS

0

Error de Cuantización y Error Topográfico

Select Splitted Dataset Part Statistics Winners Target Map

Freq. Map + Component Plans U-Matrix Anomaly Detection

Save Model

Metric	Value
Quantization Error(Total)	13186.2964
Mean Quantization Error	0.7757
Topographic Error	0.0773

Calculate

SOM Analytic Tool

Data Analysis

ORDER BY ERROR

SELECTED MODEL TO ANALYZE DATA

MQE	Topographic Error	Training Time
0.8114	0.0871	00 h 00 m 04
0.8094	0.1035	00 h 00 m 04
28707.2489	1	00 h 00 m 04
0.7279	0.878	00 h 00 m 02
1.148	0.9776	00 h 00 m 02
1.1229	0.9991	00 h 00 m 05



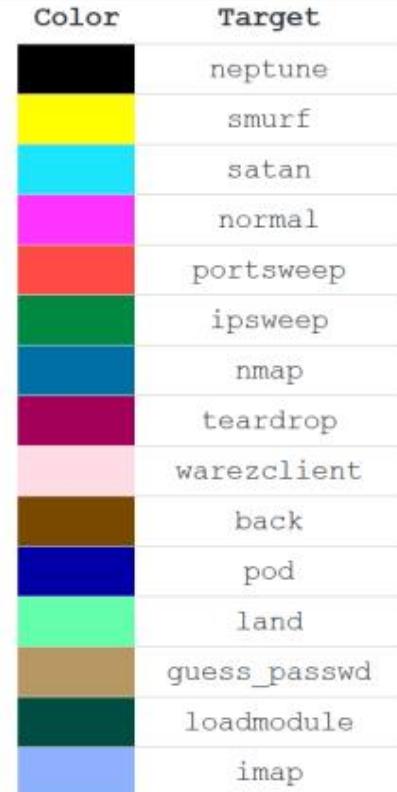
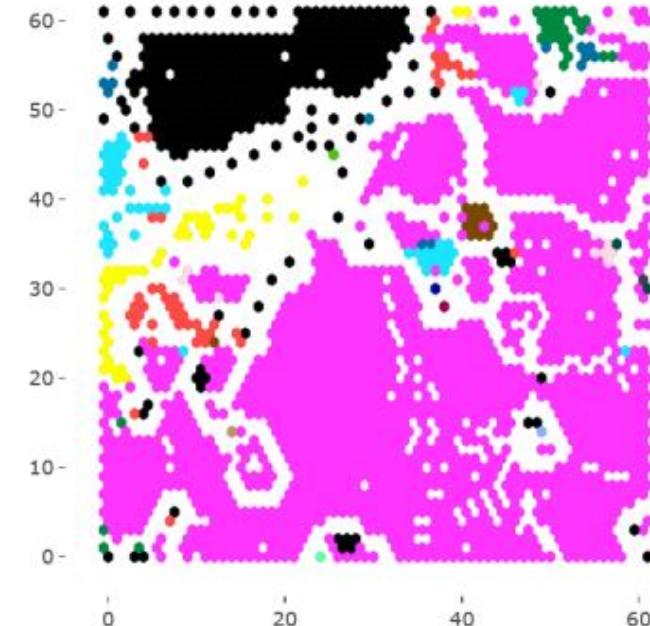
1

Mapa de Neuronas Ganadoras

DISTRIBUCIÓN DE PAÍSES POR TIPOS DE VACUNAS



ATAQUES EN RED



Cada BMU es representada por el target más frecuente en ella

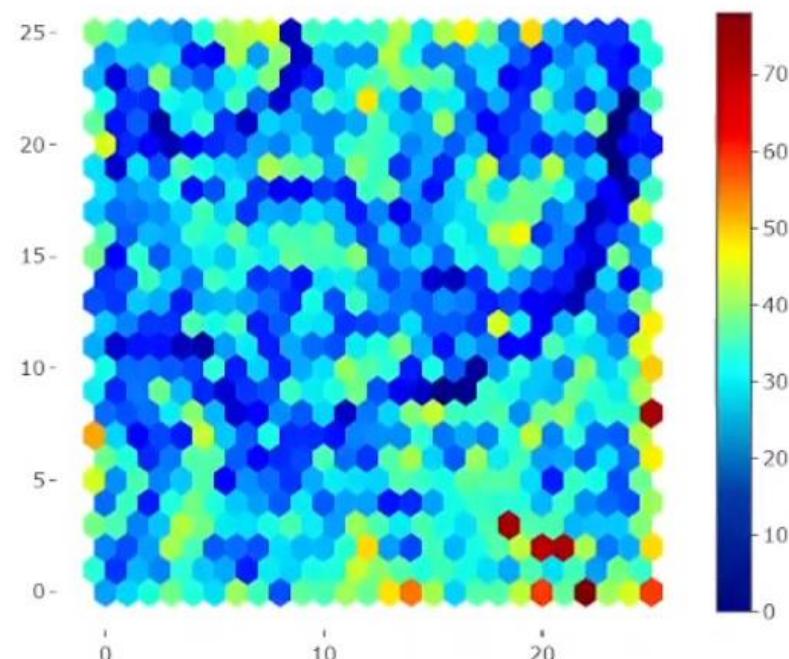


2

Mapa de Frecuencias

Frequency Map

26 x 26 neurons.



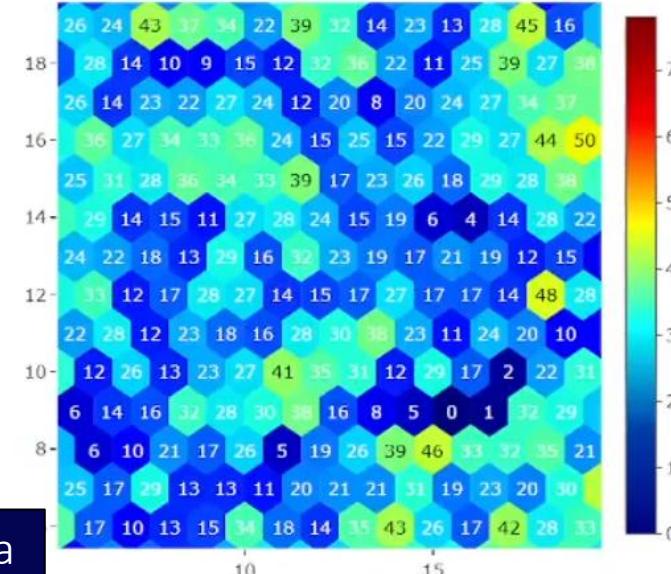
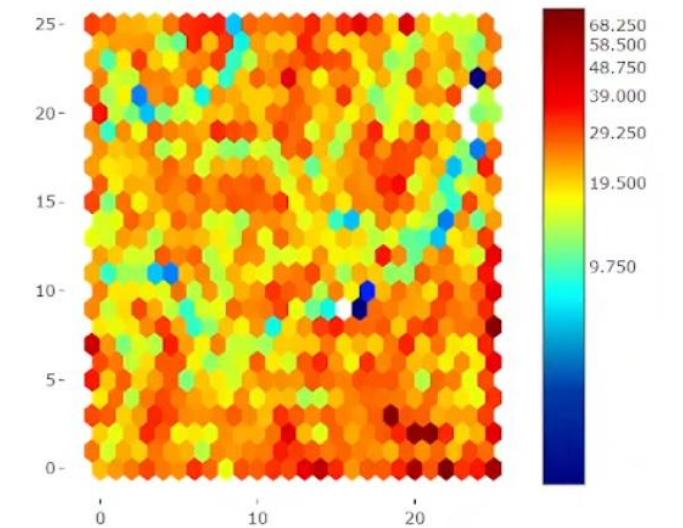
- Label Neurons
- Linear Scale
- Logarithmic Scale

ESCALA
LOGARÍTMICA

ZOOM

+

NEURONAS
ETIQUETADAS

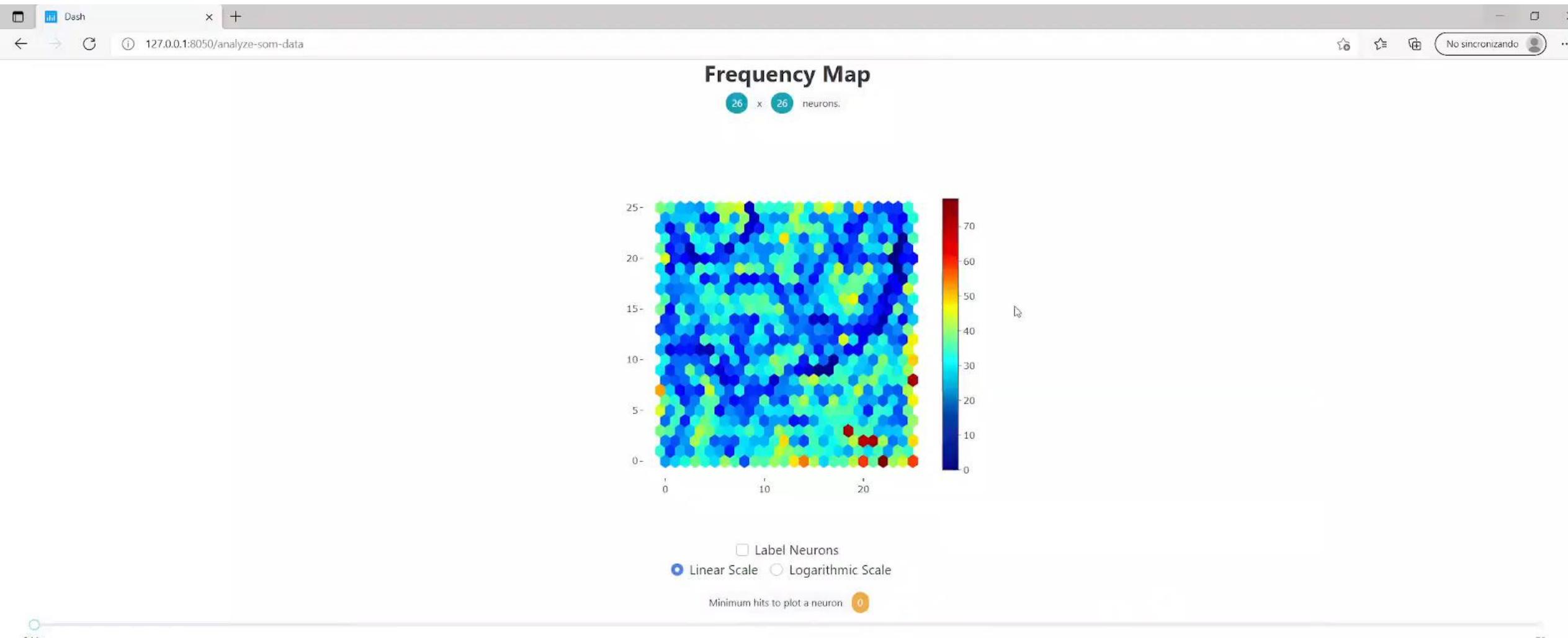


Cuantas veces ha sido nombrada BMU cada neurona

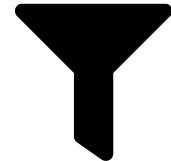


3

Mapas de Componentes

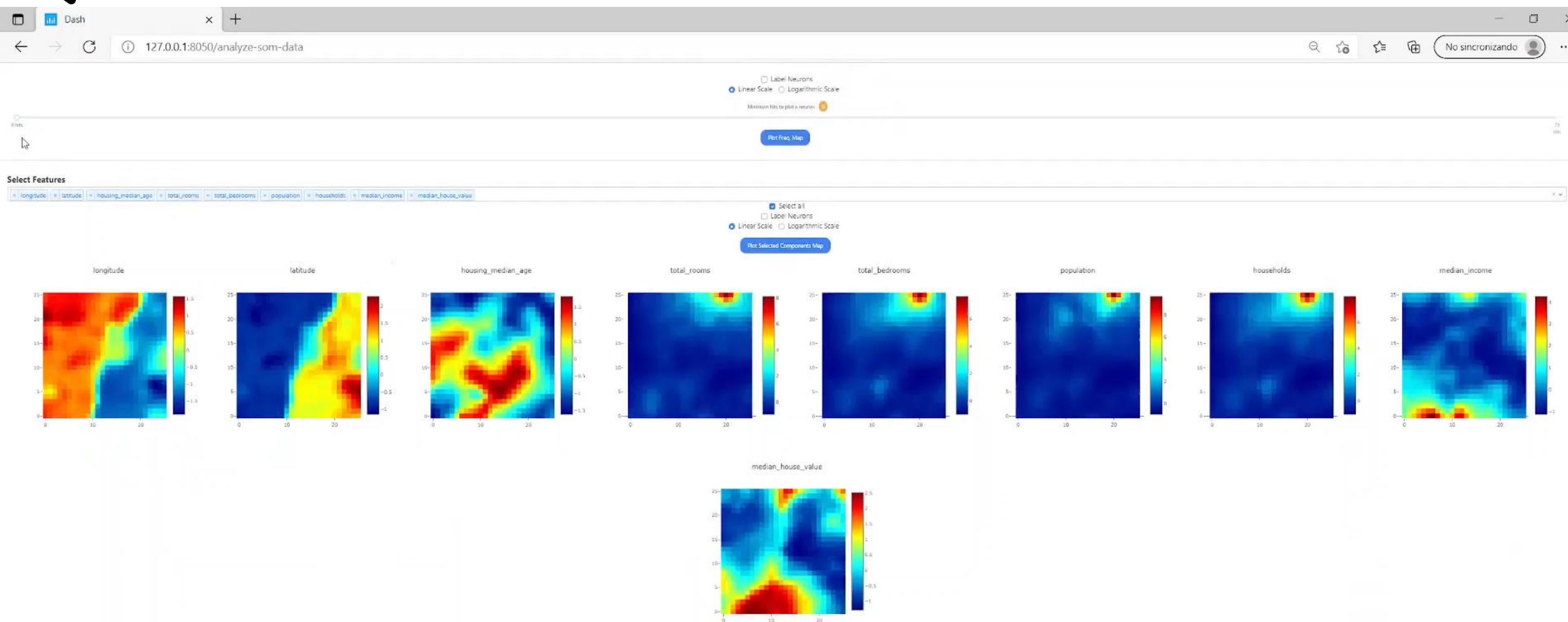


Pesos de cada característica de entrenamiento

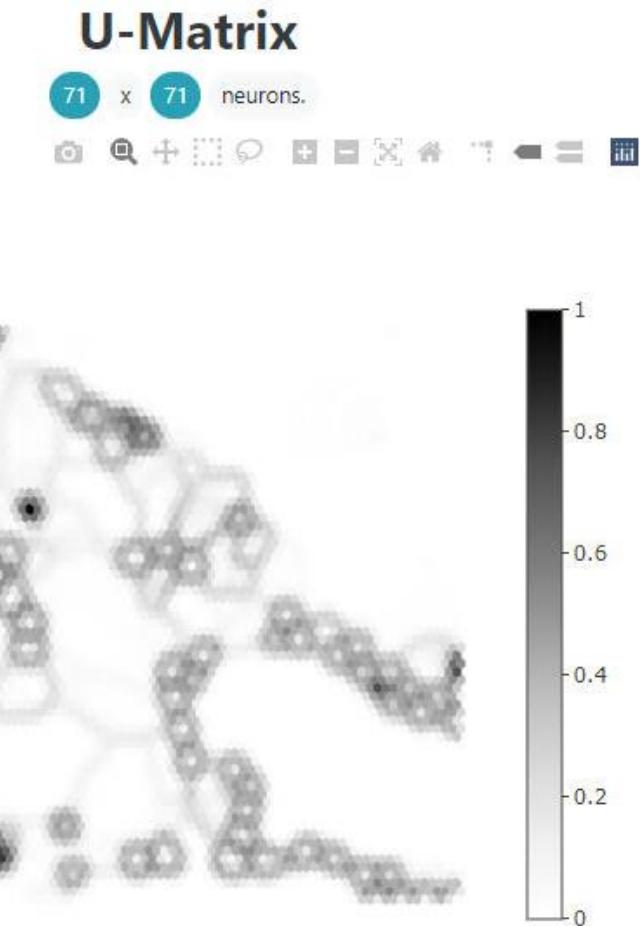
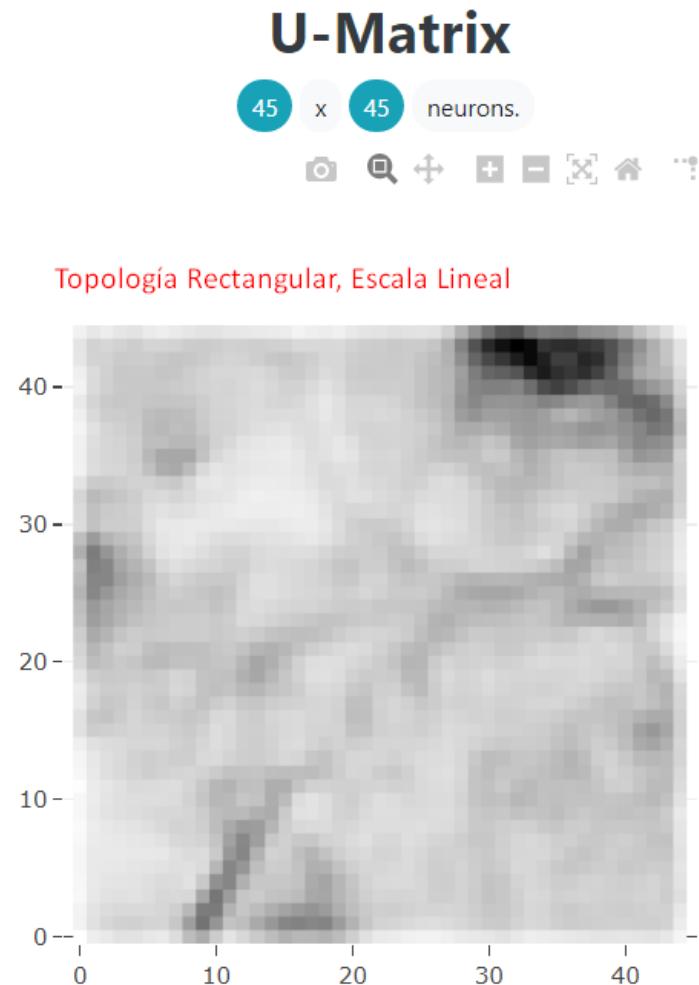


4

FILTRO Mapas de Frecuencias y Componentes



Filtro de Calidad en la Representación



Diferencia Brusca en los Pesos del Vecindario



6 Búsqueda Automática de Hiperparámetros

Dash x + 127.0.0.1:8050/train-som No sincronizando

Single Train Multi-Train Grid Hyperparameters Search Random Hyperparameters Search

Hyperparameter Selection

Params to make a search

- Grid Size Sampling Ranges** Add
- Distance Function** Add
- Weights Initialization** Added

Values to search with:

- PCA: Principal Component Analysis
- Random
- No Weight Initialization

Scoring metric to select best model

- Mean Quantization Error

Vertical Grid Size: 26
Horizontal Grid Size: 26
Distance Function: Euclidean
Learning Rate: 0,5
Map Topology: Rectangular
Neighborhood Function: Gaussian
Gaussian Sigma: 1,5
Max Iterations: 17000
Seed: Select Seed

Rand Search for Optimum Hyperparameters

Escribe aquí para buscar F1 G PDF P Google Chrome Microsoft Edge 24°C Despejado 01/07/2021 2:18

Grid Search y Rand Search



6 Búsqueda Automática de Hiperparámetros

Hyperparameter Selection

Params to make a search

- Add Grid Size Sampling Ranges
- Add Distance Function
- Added Weights Initialization

Values to search with:

- PCA: Principal Component Analysis
- Random
- No Weight Initialization

Scoring metric to select best model

Mean Quantization Error

Vertical Grid Size
26

Horizontal Grid Size
26

Distance Function
Euclidean

Learning Rate
0.5

Map Topology
Rectangular

Neighborhood Function
Gaussian

Gaussian Sigma
1.5

Max Iterations
17000

Seed
 Select Seed

Rand Search for Optimum Hyperparameters

VALIDACIÓN
CRUZADA

Parámetros Óptimos

Validation Score: Mean Quantization Error	Distance Function	Weights Initialization
0.7852 +- 0.0065	euclidean	no_init

[Add Params to MultiTrain Tab](#) [Discard HyperParameters](#) [Analyze Model with current Hyperparameters](#)

Grid Search y Rand Search



7

Detector de Anomalías con SOM's

SOM Analytic Tool

Data Analysis

Show Model Info/ Change Model Selection

Select Splitted Dataset Part Statistics Winners Target Map Freq. Map + Component Plans U-Matrix Anomaly Detection Save Model

Show Tutorial

Upload File to Search for Anomalies in Data

Drag and Drop or Click to Select File (.csv)

Take as Anomaly, data that has similarity percentage smaller than Small values increasing from 0 recommended, for an optimum search

means All Data will be Classified as Normal
 means All Data will be Classified as Anomaly

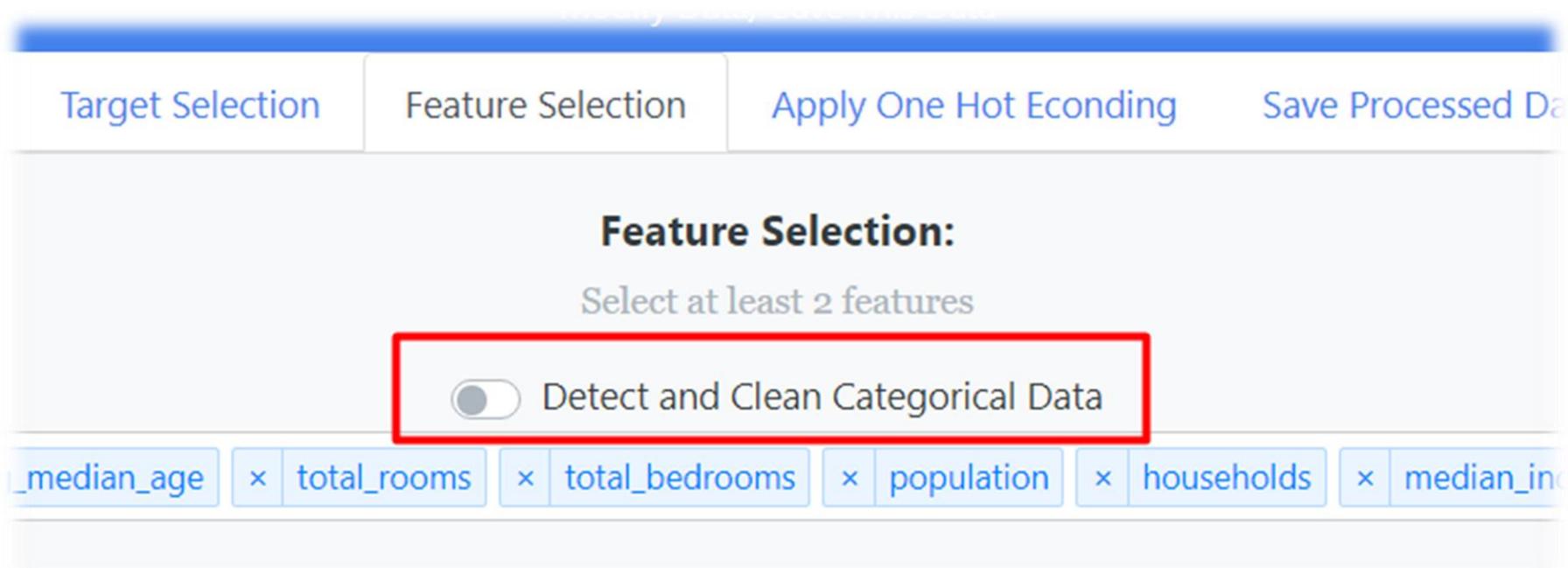
Search for Anomalies

Datos Potencialmente Anómalos + Ranking Atributos Anómalos

A

8

Auto-detectar datos categóricos durante el Pre-Procesado dato



FILTRO:En datos numéricos autodetecta datos categóricos



9

MULTI Entrenamiento/Análisis de SOM's

Multi - Entrenamiento

Hyperparameter Selection

	Hor. Size	Ver. Size	Learning Rate	Neighborhood Function	Distance Function	Gaussian Sigma	Max
—	x 26	26	0.5	gaussian	euclidean	1.5	
—	x 26	26	0.7	bubble	cosine	1	
—	x 20	20	1.2	mexican_hat	manhattan	1	
—	x 15	15	1.2	triangle	chebyshev	0.999999	

Multi - Análisis

SOM Analytic Tool

Data Analysis

ORDER BY ERROR

	MQE	Topographic Error	Training Time	Hor. Size	Ver. Size	Learning Rate	Neighborhood
○	0.8114	0.0871	00 h 00 m 04 s	24	24	0.5	gaussian
○	0.8094	0.1035	00 h 00 m 04 s	24	24	0.5	gaussian
○	28707.2489	1	00 h 00 m 04 s	20	20	0.7	mexican_hat
○	0.7279	0.878	00 h 00 m 02 s	20	20	0.9	triangle
○	1.148	0.9776	00 h 00 m 02 s	15	15	0.2	bubble
○	1.1229	0.9991	00 h 00 m 05 s	30	30	0.2058	mexican_hat

Replot Graphs

SELECTED MODEL TO ANALYZE DATA



10 Multi Targets

Select Splitted Dataset Part Statistics **Winners Target Map**

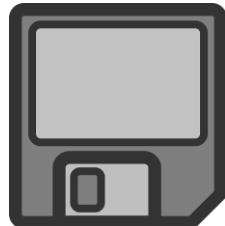
Freq. Map + Component Plans U-Matrix Anomaly Detection

[Save Model](#)

Target not selected yet !
Please select a target below to print winners map.

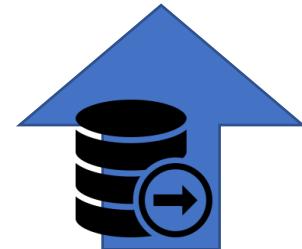
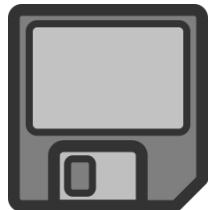
Select...

- rank
- country
- category
- civil_liberties



11

Carga/Guardado de Modelos



Saved Models

california_hex_som.pickle

Model Vector Dimensionality: 8

Model Trained With Features:

longitude latitude housing_median_age total_rooms total_bedrooms population households median_income

Load Selected Model

INFORMACIÓN
DE
MODELOS GUARDADOS

Muestra Información Sobre Los Atributos de Entrenamiento



12

Guardado de Datos Procesados

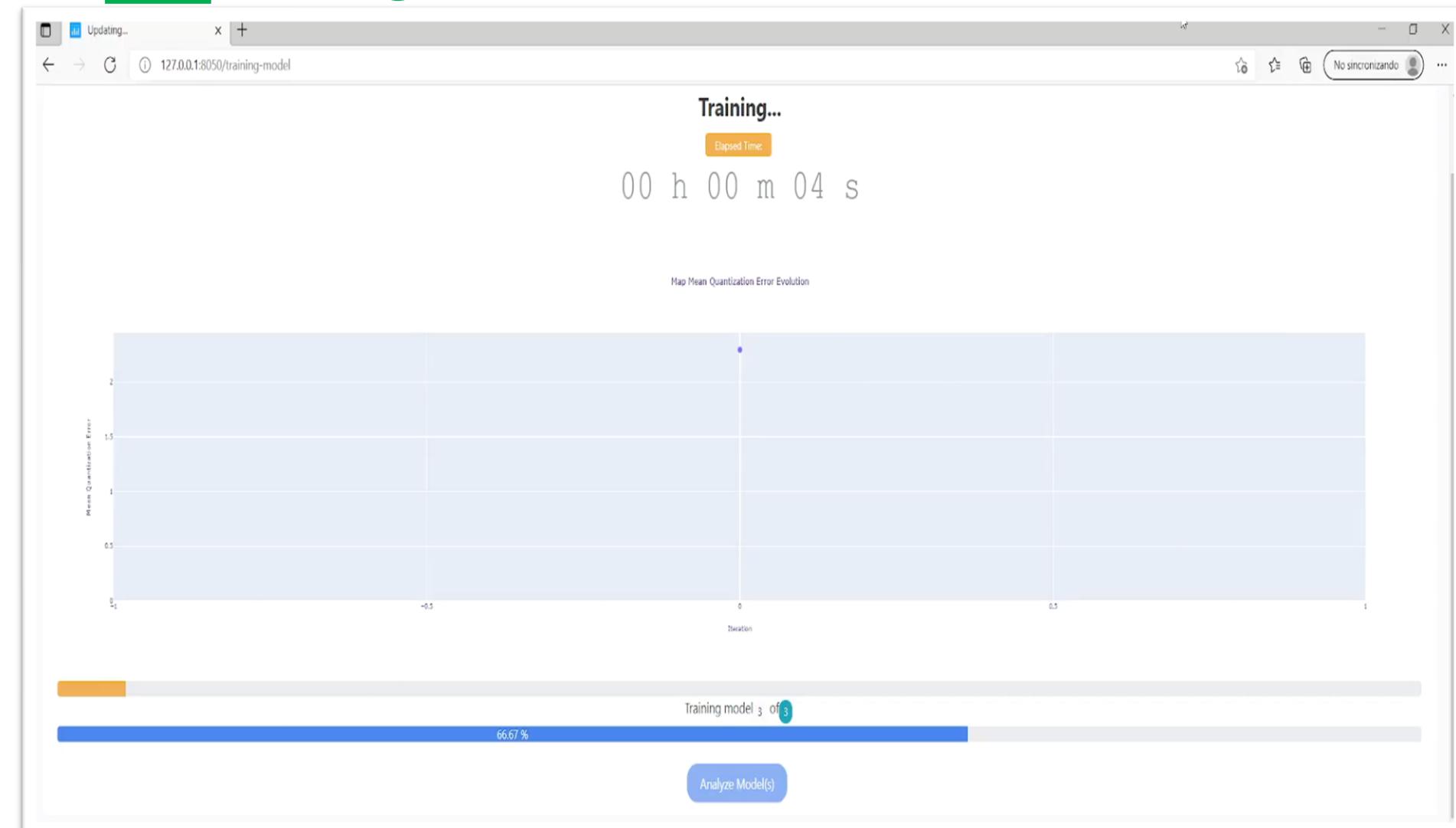
The screenshot shows a web application interface for data processing. At the top, there is a header bar with a back button, forward button, refresh button, and a URL field showing "127.0.0.1:8050". To the right of the URL are several icons: a star, a sync icon, a user icon, and a more options icon. Below the header is a table displaying a dataset with 10 rows and 11 columns. The columns represent various attributes: 0, 0, 1, 0, -122.23, 37.88, 41, 880, 129, 1, and 2. A blue button at the bottom of the table row says "Modify Data/ Save This Data". Below the table is a navigation bar with tabs: "Dataset Size", "Target Selection", "Feature Selection", "Apply One Hot Encoding", and "Save Processed Data to .csv File". The "Feature Selection" tab is currently active. A sub-section titled "Feature Selection:" contains the instruction "Select at least 2 features" and a "Detect and Clean Categorical Data" checkbox. Below this are several categorical feature names listed in a row: ocean_proximity_INLAND, ocean_proximity_ISLAND, ocean_proximity_NEAR BAY, ocean_proximity_NEAR OCEAN, longitude, latitude, housing_median_age, total_rooms, total_bedrooms, households, and median_income. At the bottom of the page are two buttons: "Train New Model" and "Load Pre-Trained Model". The bottom left corner of the page shows the URL "127.0.0.1:8050/#".

Muestra Información Sobre Los Atributos de Entrenamiento

Err

13

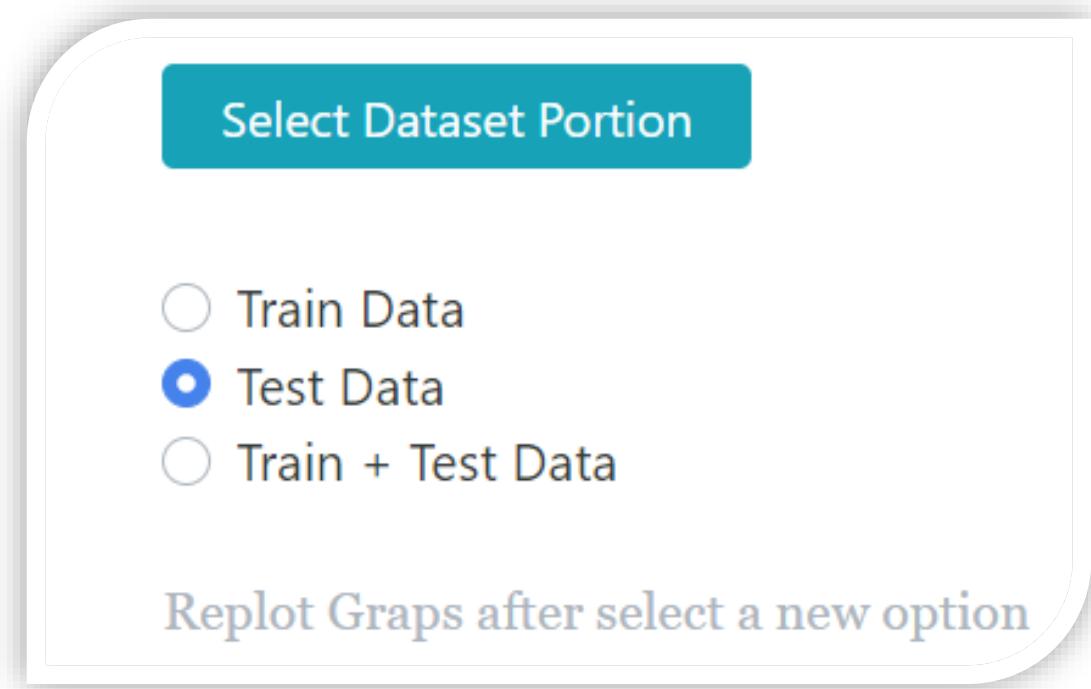
Progreso del Error



Entrenamientos
permiten uso de
semilla

Durante el entrenamiento

14 Seleccionar Porción de Datos para Análisis



Útil cuando los datos disponibles son limitados

```
-->Shuffling Data...
-->Shuffling Complete.
-->Standardizing Train Data...
-->Standardizing Complete.
-->Training SOM 0 ...
-->Training Complete!
    Elapsed Time: 4.7293455600738525 seconds
```

Info. Detallada por Terminal

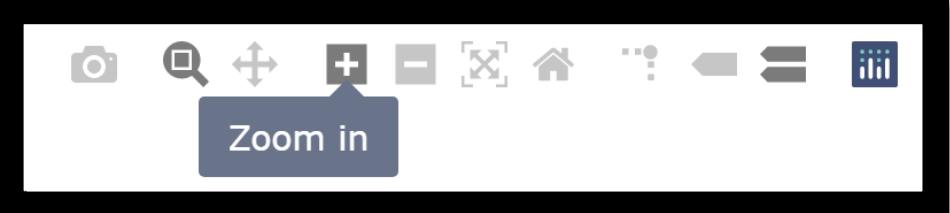
```
#PLOT CONF
DEFAULT_HEATMAP_PX_HEIGHT = 500
DEFAULT_HEATMAP_PX_WIDTH = 500
#DEFAULT_HEATMAP_COLORSCALE = 'Viridis'
DEFAULT_HEATMAP_COLORSCALE = 'Jet'
UMATRIX_HEATMAP_COLORSCALE = 'Greys'

DEFAULT_DATASET_ROWS_PREVIEW = 10
CATEGORICAL_TABLE_LEGEND_ELEMENTS_PER_PAGE = 15
```

Ajustes para desarrolladores en Config.py

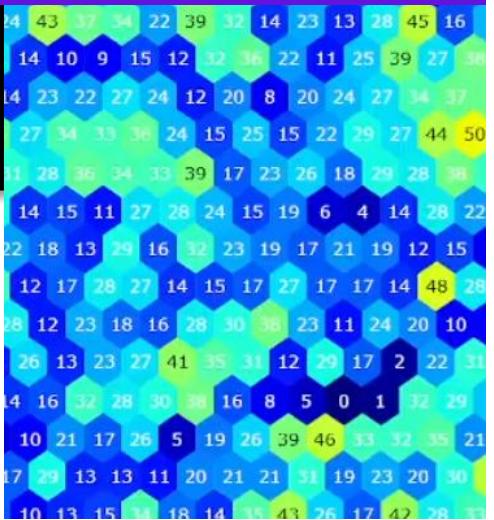


Barra Plotly



- Capturas
- Zoom
- Marcas
- ...

Etiquetado en Neuronas



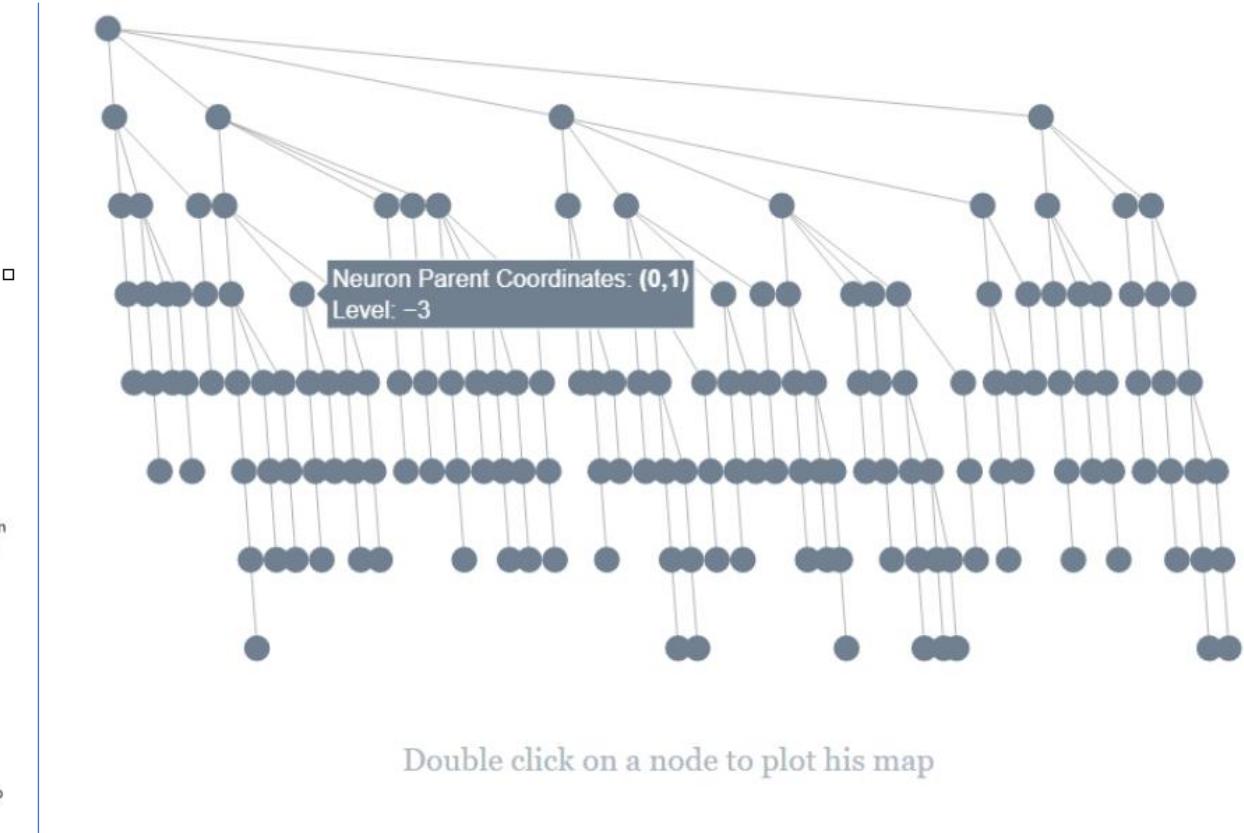
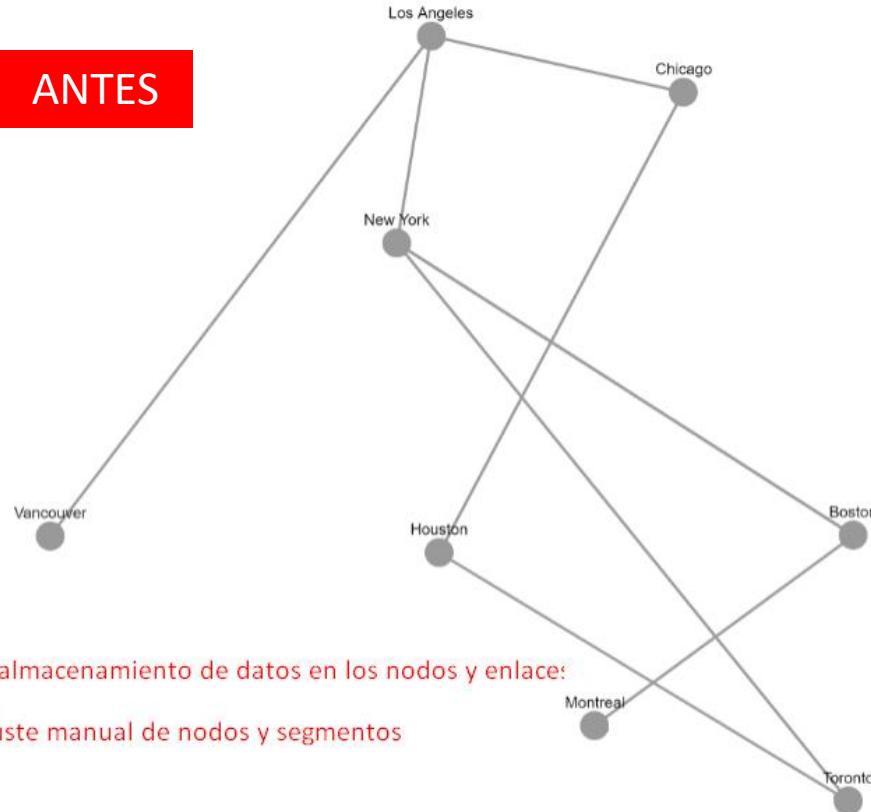
Escalas

Linear Scale Logarithmic Scale

IMPLEMENTACIONES

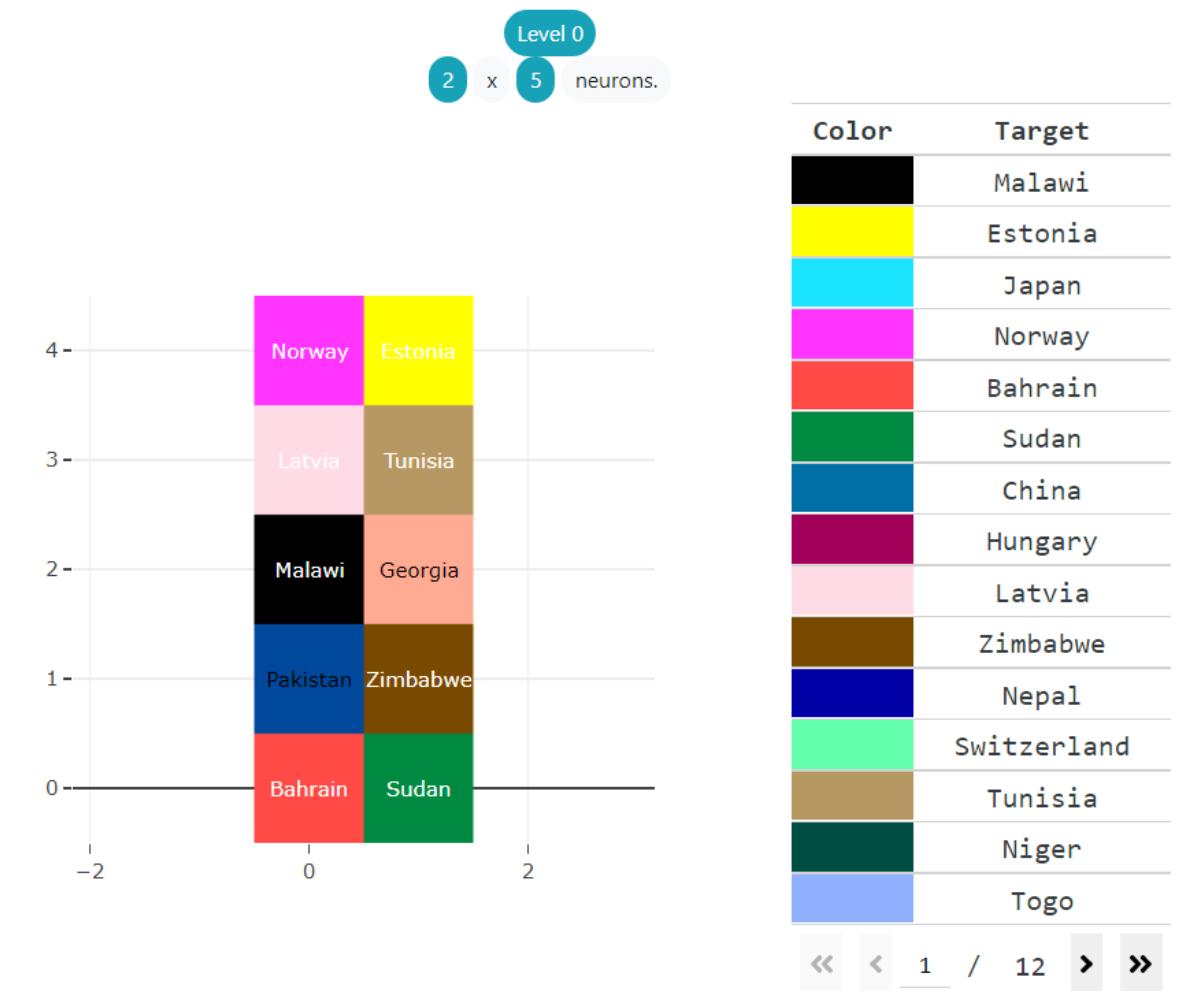
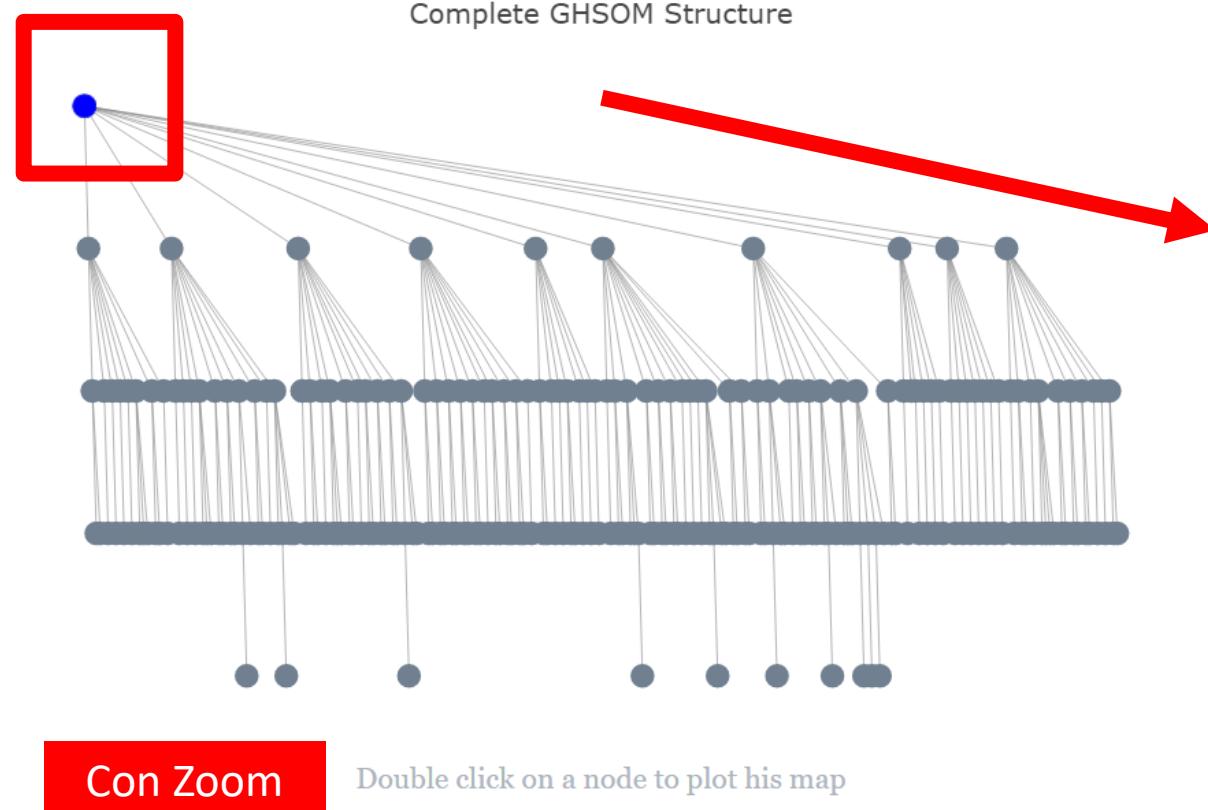


ANTES



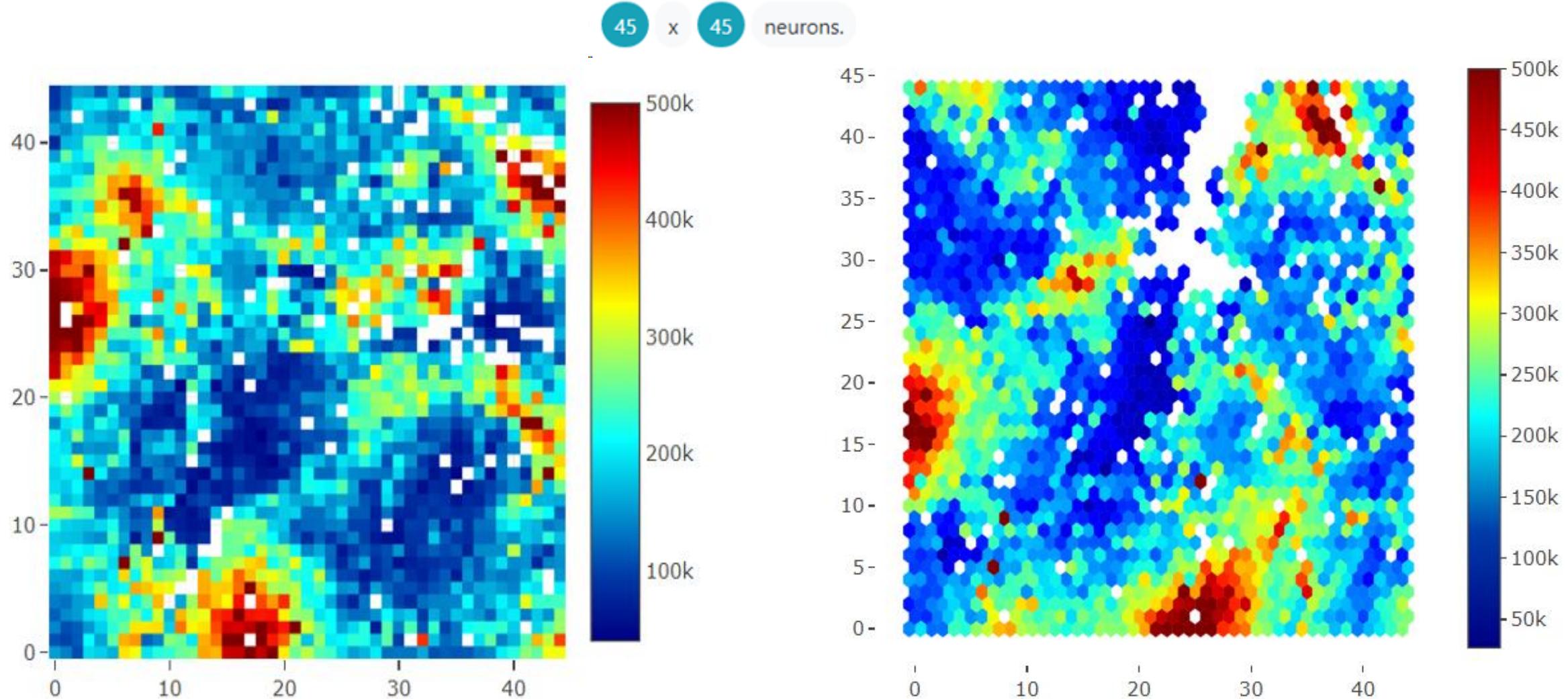
Árbol Interactivo en Dash

Winners Target per Neuron Map

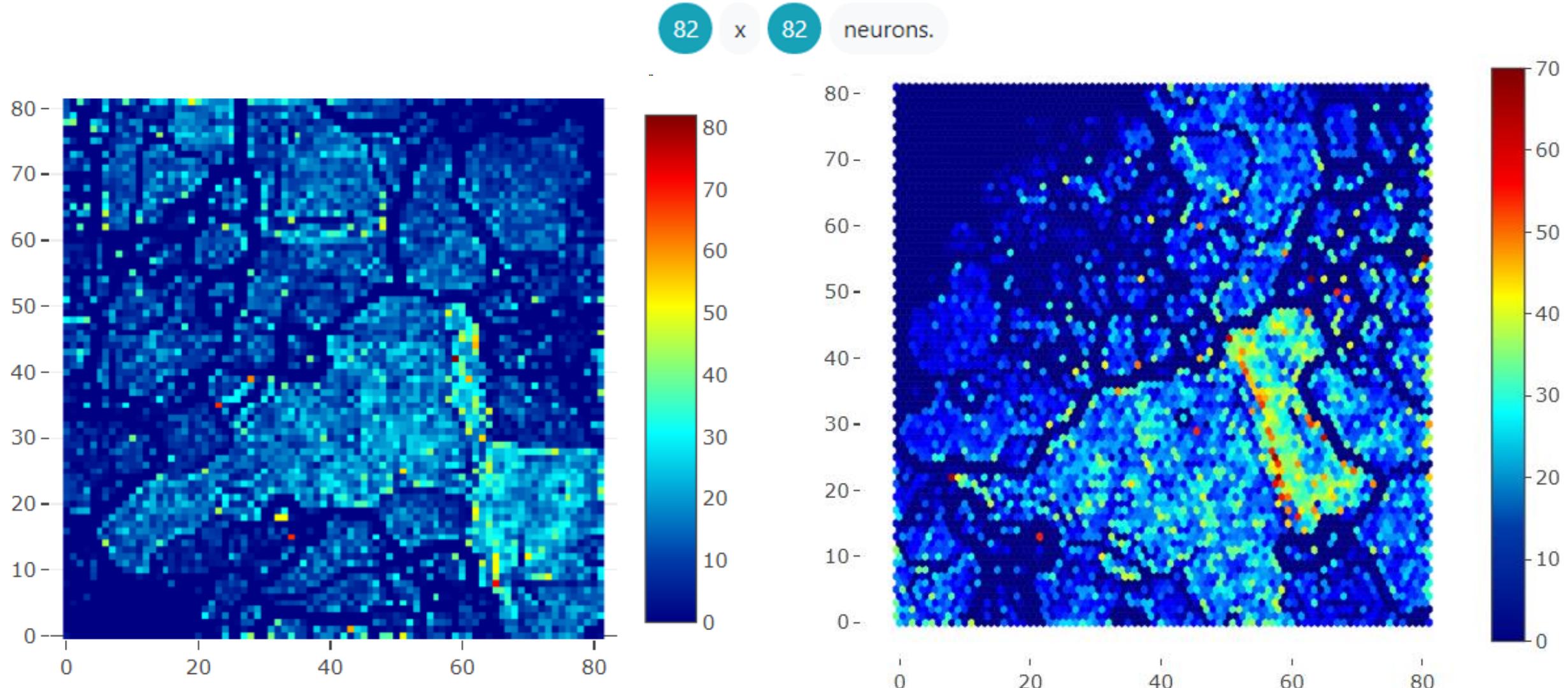


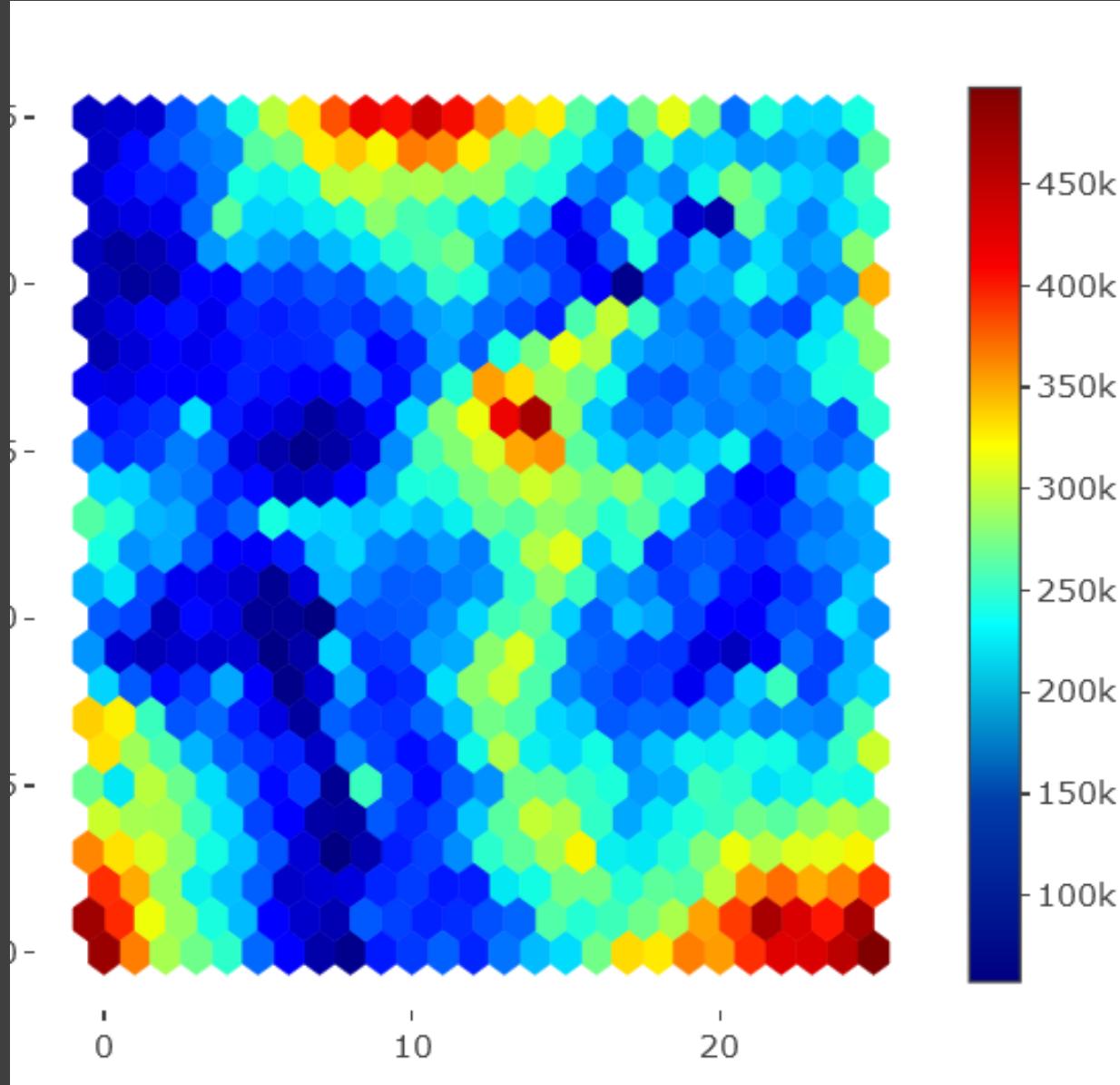
Implementación exclusiva para la aplicación

TOPOLOGÍA HEXAGONAL



TOPOLOGÍA HEXAGONAL



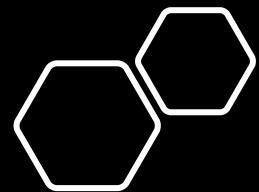


GRID HEXAGONAL PYTHON

Implementación
exclusiva para la
aplicación

Error Topográfico

IMPLEMENTACIÓN
PARA
SOM'S
HEXAGONALES



TAMAÑO
AUTOMÁTICO
RECOMENDADO

Single Train

Multi-Train

Grid Hyperparameters Se

Vertical Grid Size

26

Horizontal Grid Size

26

Distance Function



Hasta **269** Targets
Catégoricos con
Colores
Visualmente
Diferenciables

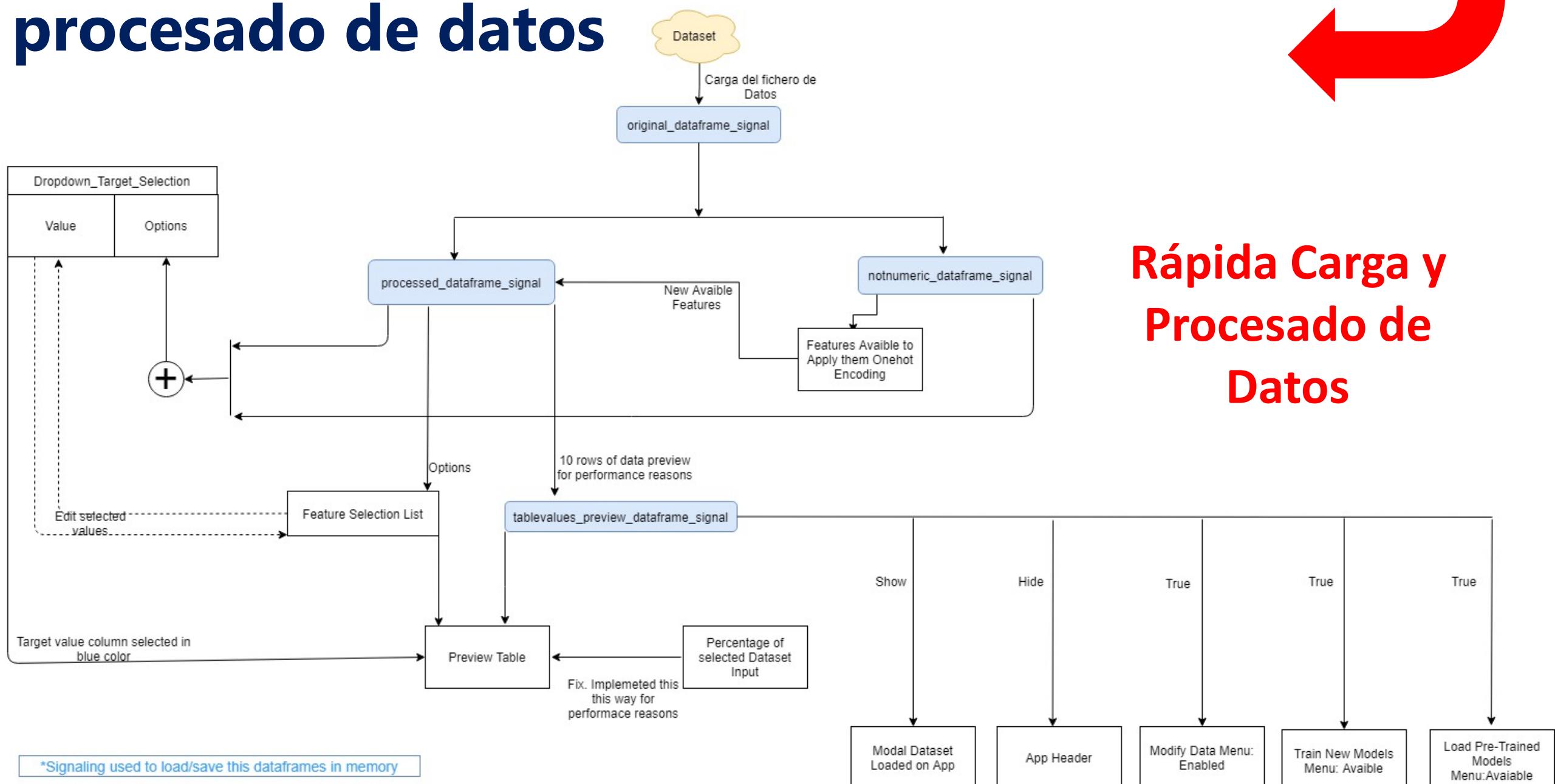
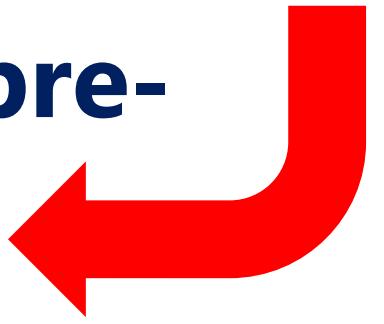
Color	Target
Black	Latvia
Yellow	Pakistan
Cyan	Canada
Magenta	Hungary
Red	Bulgaria
Green	France
Blue	Slovakia
Dark Purple	Tanzania
Light Pink	Comoros
Brown	Democratic Republic of the Congo
Dark Blue	Kazakhstan
Cyan	Ecuador
Brown	Peru
Dark Green	Burundi
Light Blue	Nicaragua

POR DEFECTO: 30

269

Color	Target
Black	Latvia
Yellow	Pakistan
Cyan	Canada
Magenta	Hungary
Red	Bulgaria
Green	France
Blue	Slovakia
Purple	Tanzania
Light Blue	Comoros
Brown	Democratic Republic of the Congo
Dark Blue	Kazakhstan
Light Green	Ecuador
Khaki	Peru
Dark Green	Burundi
Light Blue	Nicaragua
« « 1 / 12 » »	
Color	Target
Grey	Namibia
Dark Red	Portugal
Grey	Germany
Light Yellow	Palestine
Dark Green	Ireland
Green	United Arab Emirates
Blue	Netherlands
Dark Grey	Ghana
Red	Fiji
Dark Grey	Ivory Coast
Red	Afghanistan
Khaki	Jordan
Teal	Luxembourg
Orange	Argentina
Pink	Guyana
« < 2 / 12 > »	
Color	Target
Magenta	Burkina Faso
Orange	Bolivia
Light Blue	Iraq
Dark Blue	Yemen
Grey	Guatemala
Green	Cuba
Dark Purple	Bosnia and Herzegovina
Blue	Honduras
Dark Teal	Japan
Teal	Morocco
Dark Brown	Switzerland
Yellow	Kyrgyzstan
Light Cyan	Costa Rica
Light Purple	Estonia
Red	Vietnam
« < 3 / 12 > »	
Color	Target
Grey	North Macedonia
Dark Red	Laos
Light Cyan	Russia
Orange	Austria
Light Teal	Eritrea
Light Brown	India
Green	Turkmenistan
Dark Green	Libya
Dark Blue	Albania
Light Blue	Georgia
Yellow	Cameroon
Dark Orange	Greece
Light Teal	Cambodia
Light Brown	Montenegro
Dark Grey	Mauritania
Dark Green	Hong Kong
« < 9 / 12 > »	
Color	Target
Light Green	Jamaica
Light Green	Philippines
Dark Green	Slovenia
Dark Blue	Belgium
Dark Purple	Sierra Leone
Light Cyan	Haiti
Light Purple	Senegal
Dark Teal	Togo
Light Brown	Ethiopia
Light Teal	Uzbekistan
Light Brown	Mali
Dark Purple	Paraguay
Light Orange	Republic of the Congo
Light Red	Costa Rica
Light Purple	North Korea
Light Red	Belarus
« < 4 / 12 > »	
Color	Target
Light Green	Kenya
Dark Green	Czech Republic
Dark Blue	Norway
Light Brown	Oman
Dark Grey	Malta
Light Purple	Iran
Light Teal	Gambia
Light Brown	Chile
Light Teal	Singapore
Light Green	Israel
Light Orange	Nigeria
Light Brown	Iceland
Light Green	Lithuania
Light Teal	Rwanda
Light Red	Malawi
« < 5 / 12 > »	
Color	Target
Light Green	Dominican Republic
Light Red	El Salvador
« < 12 / 12 > »	
180	

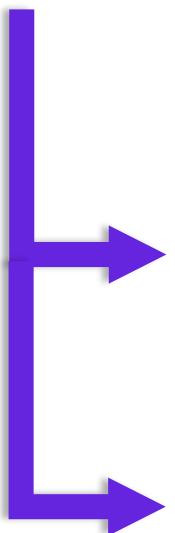
Flujo de las principales componentes del pre-procesado de datos



INTERFAZ SOM para Scikit-learn



Implementada para Grid Search y Random Search



Reutilizable para futuros usos

Implementación actual extremadamente reciente
y básica

CASOS DE ESTUDIO

GHSOM

índice de

Democracia en

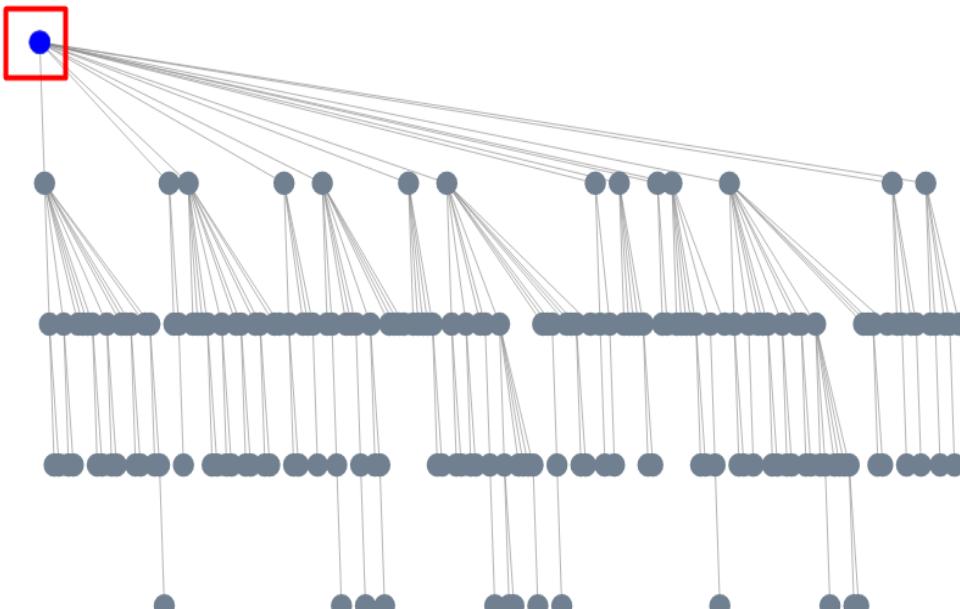
los Países del

Mundo

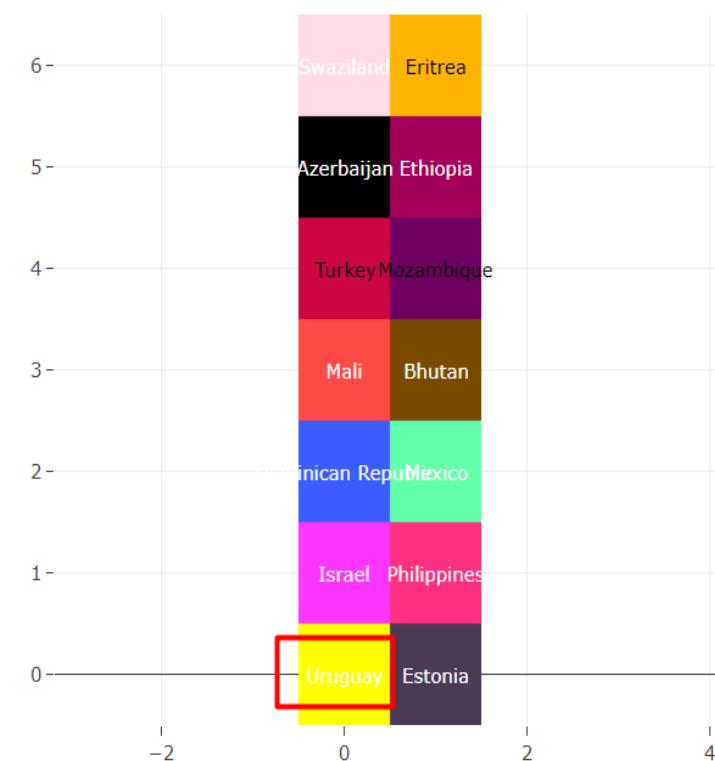
Winners Target per Neuron Map

Level 0
2 x 7 neurons.

Complete GHSOM Structure

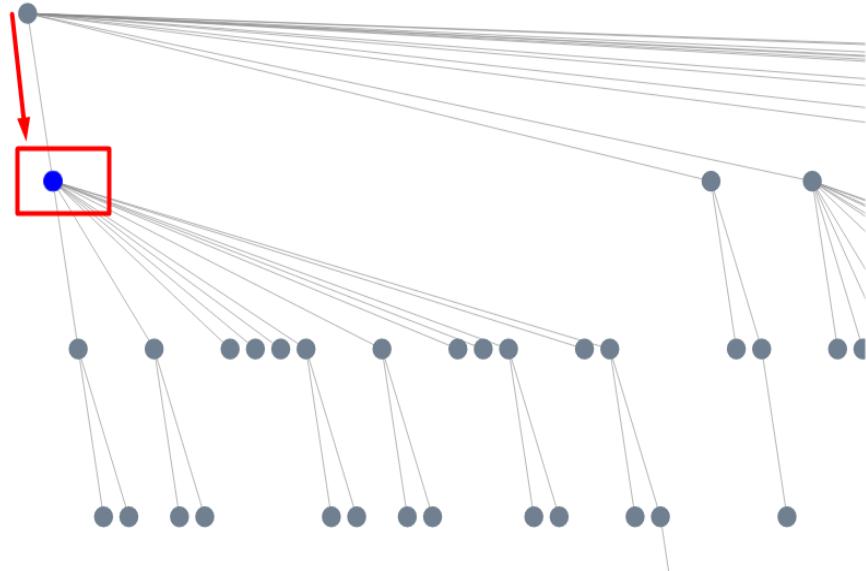


Double click on a node to plot his map

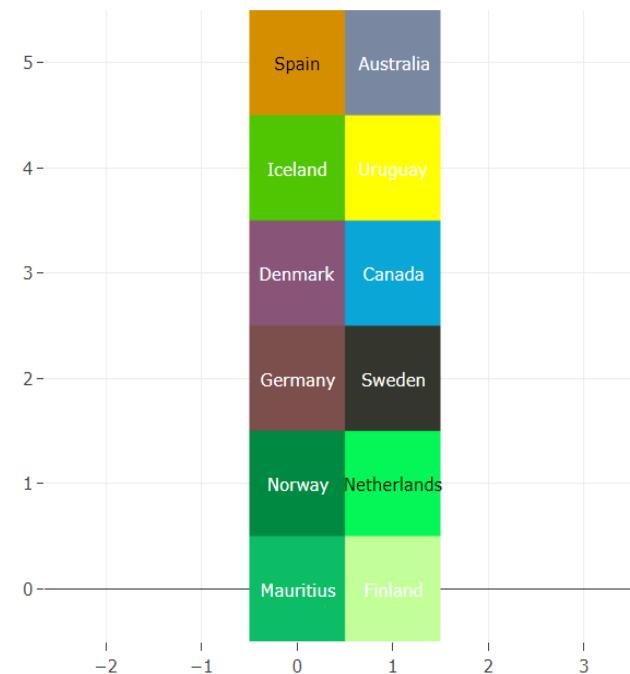


Color	Target
Black	Azerbaijan
Yellow	Uruguay
Cyan	North Korea
Magenta	Israel
Red	Mali
Green	Norway
Blue	Republic of China (Taiwan)
Purple	Ethiopia
Pink	Swaziland
Brown	Bhutan
Dark Blue	Sudan
Light Green	Mexico
Olive	Vietnam
Dark Green	Rwanda
Light Blue	Algeria

Complete GHSOM Structure

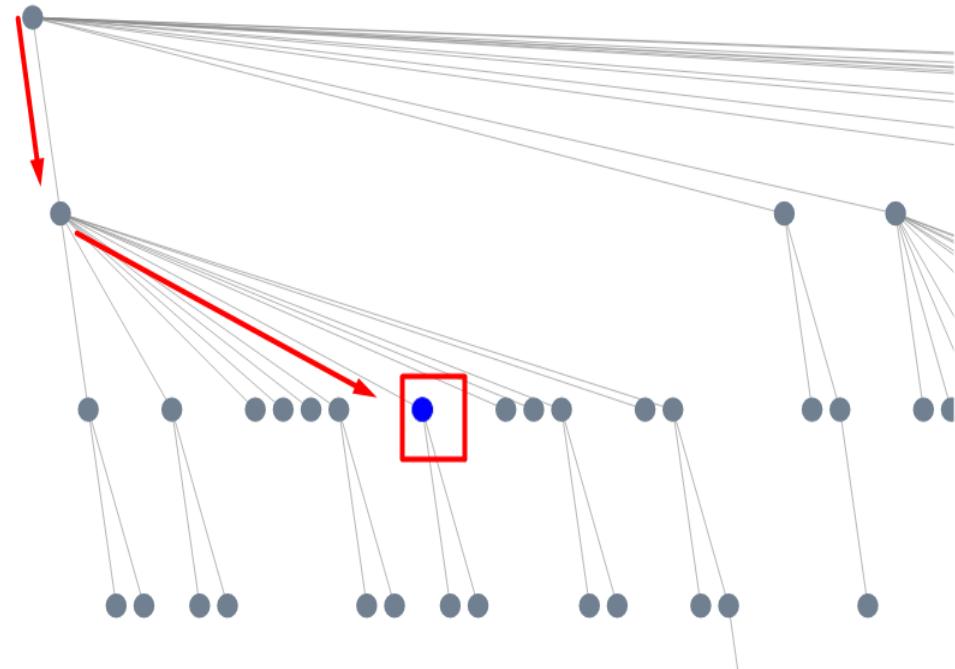


Level 1
Parent Neuron: (0,0)
2 x 6 neurons.



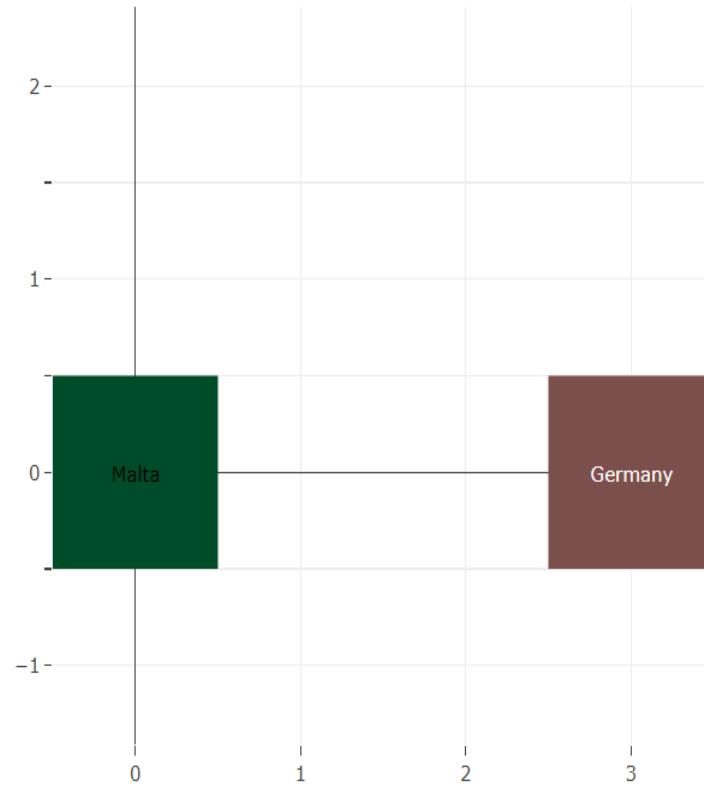
Color	Target
Black	Azerbaijan
Yellow	Uruguay
Cyan	North Korea
Magenta	Israel
Red	Mali
Green	Norway
Blue	Republic of China (Taiwan)
Purple	Ethiopia
Pink	Swaziland
Brown	Bhutan
Dark Blue	Sudan
Cyan	Mexico
Light Brown	Vietnam
Dark Green	Rwanda
Light Blue	Algeria

Complete GHSOM Structure



Double click on a node to plot his map

4 X 2 neurons.

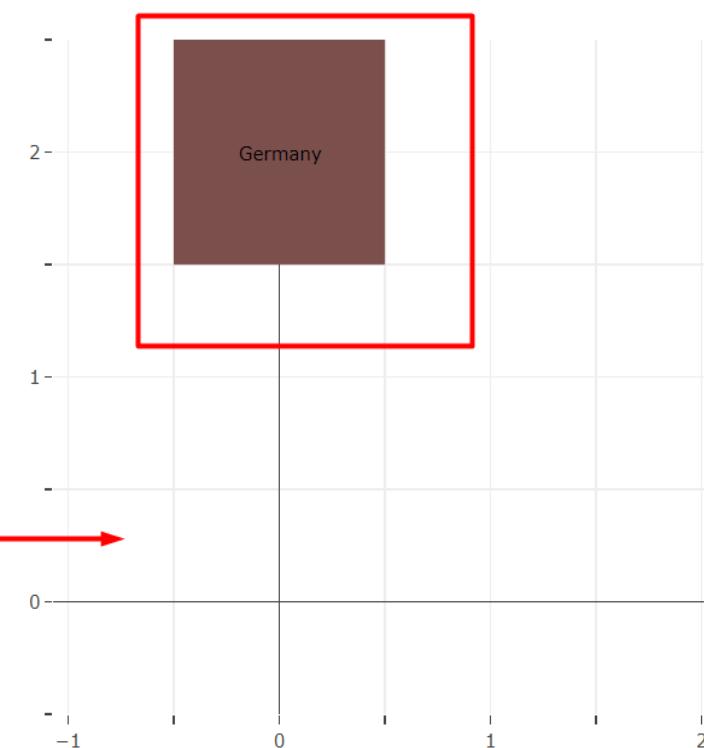
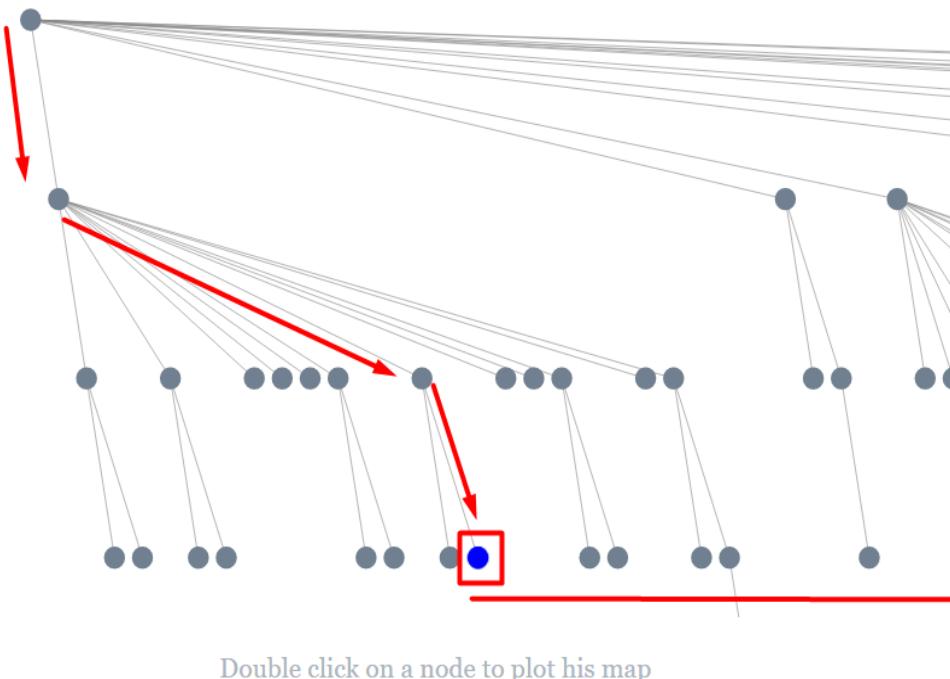


Color	Target
Black	Azerbaijan
Yellow	Uruguay
Cyan	North Korea
Magenta	Israel
Red	Mali
Green	Norway
Blue	Republic of China (Taiwan)
Purple	Ethiopia
Pink	Swaziland
Brown	Bhutan
Dark Blue	Sudan
Cyan	Mexico
Brown	Vietnam
Dark Green	Rwanda
Light Blue	Algeria

Winners Target per Neuron Map

Level 3
Parent Neuron: (3,0)
2 x 3 neurons.

Complete GHSOM Structure



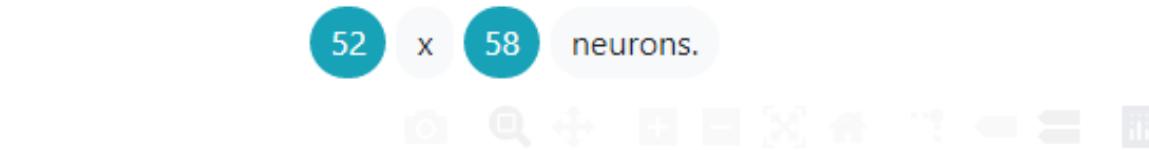
Color	Target
Black	Azerbaijan
Yellow	Uruguay
Cyan	North Korea
Magenta	Israel
Red	Mali
Green	Norway
Blue	Republic of China (Taiwan)
Purple	Ethiopia
Pink	Swaziland
Brown	Bhutan
Dark Blue	Sudan
Light Green	Mexico
Tan	Vietnam
Dark Green	Rwanda
Light Blue	Algeria

Navigation controls: << < 1 / 12 > >>

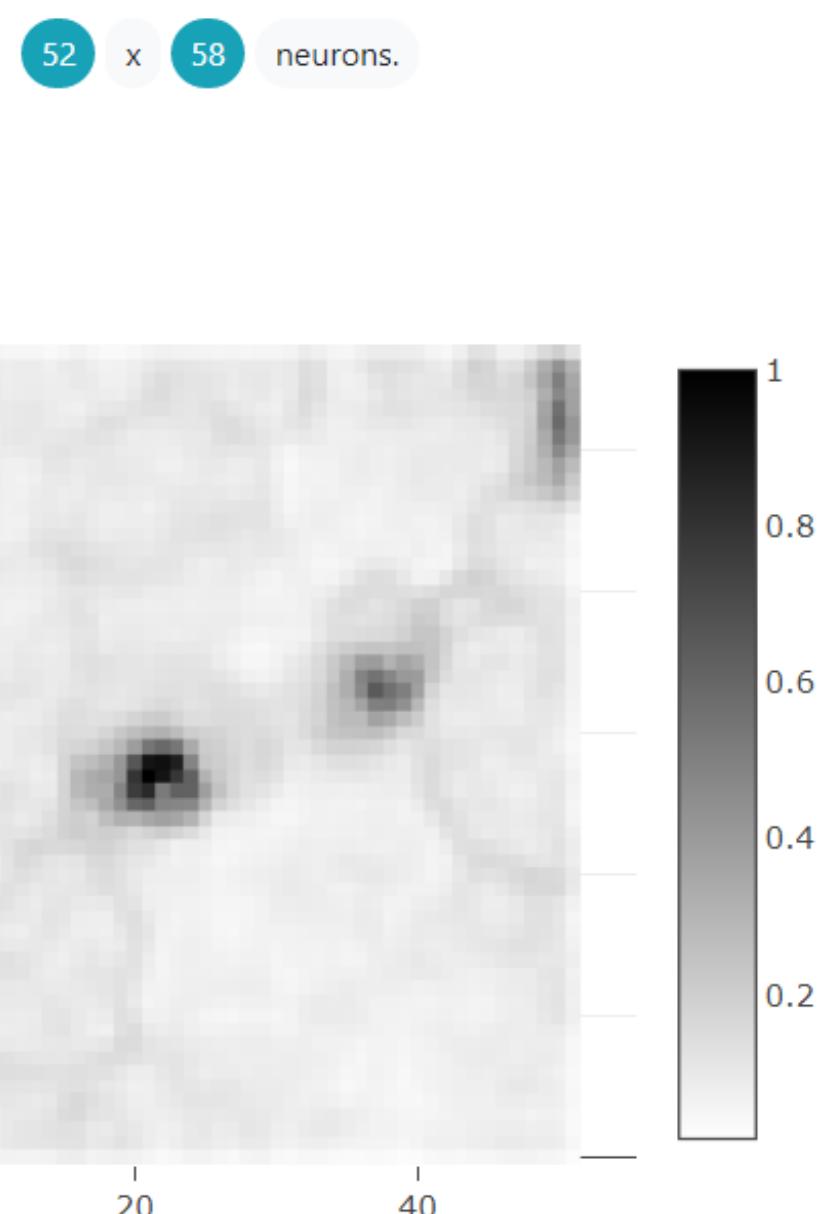
SOM Rectangular

Fraude en Tarjetas de Crédito

Winners Target per Neuron



U-Matrix



SOM Hexagonal

**Precio de la
Vivienda
En California**

45

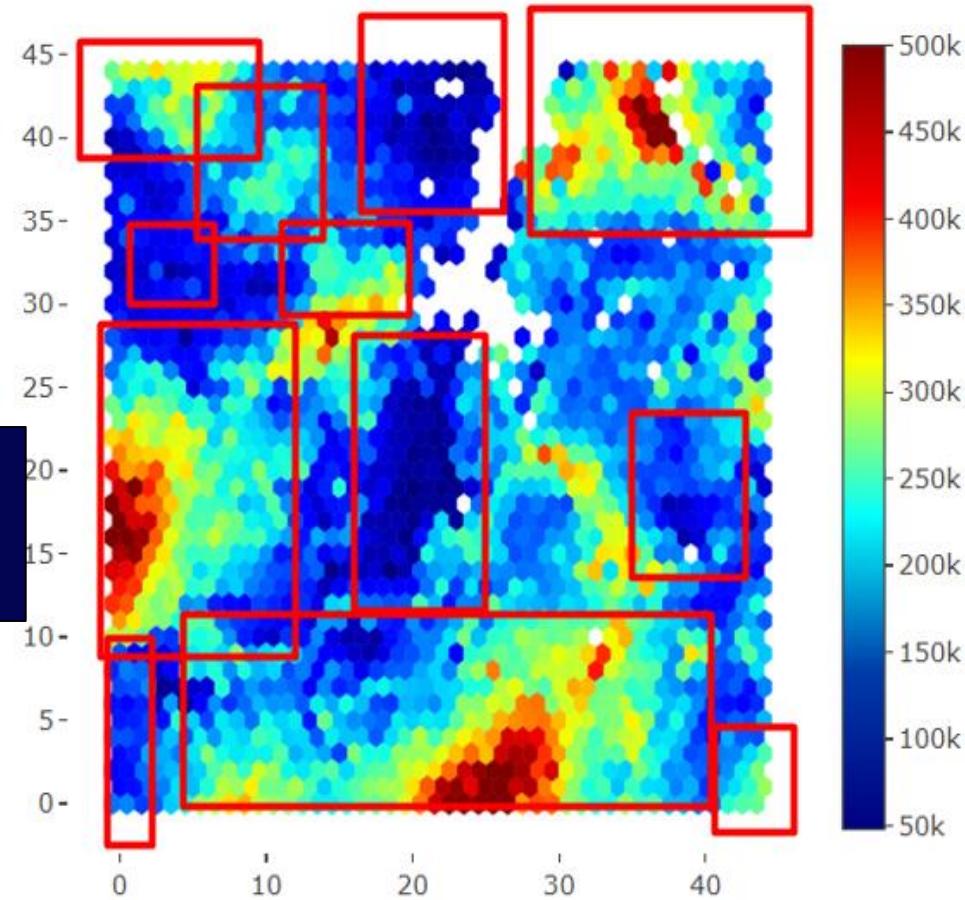
x

45

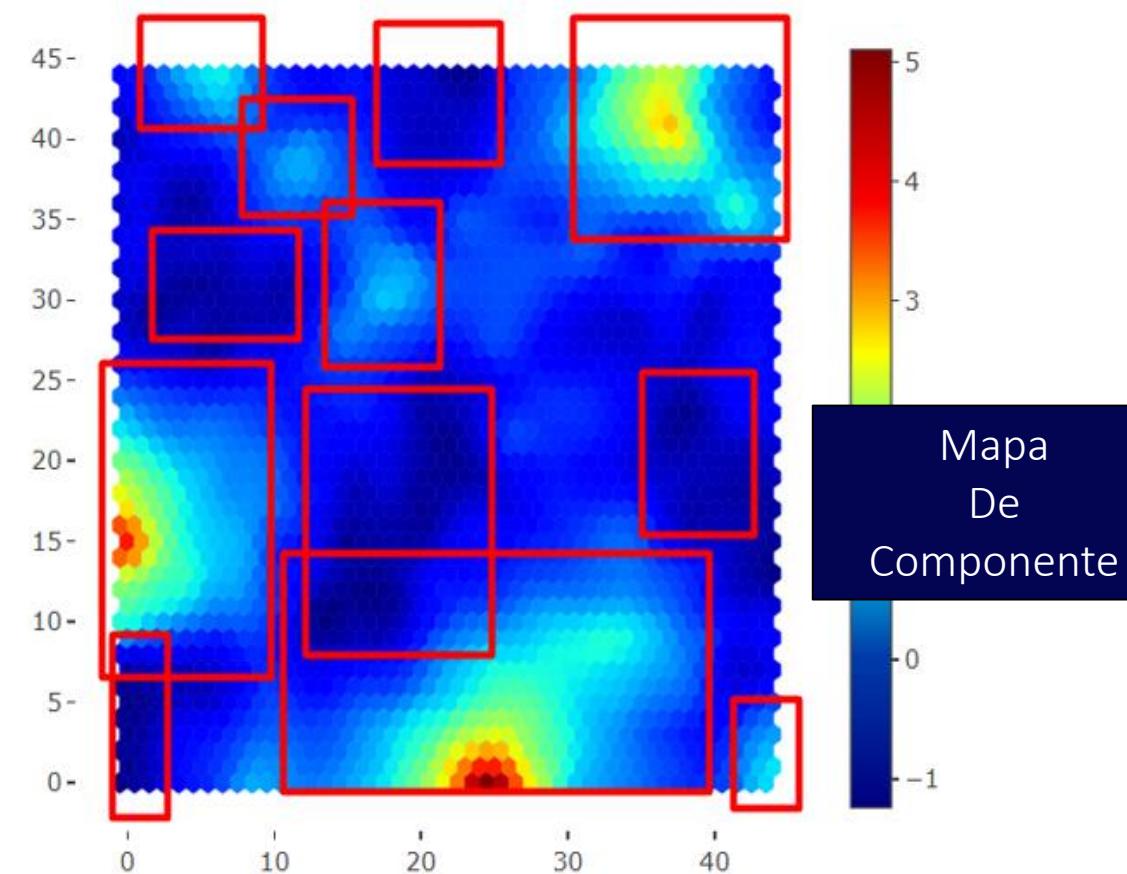
neurons.

median_income

Mapa
Neuronas
Ganadoras

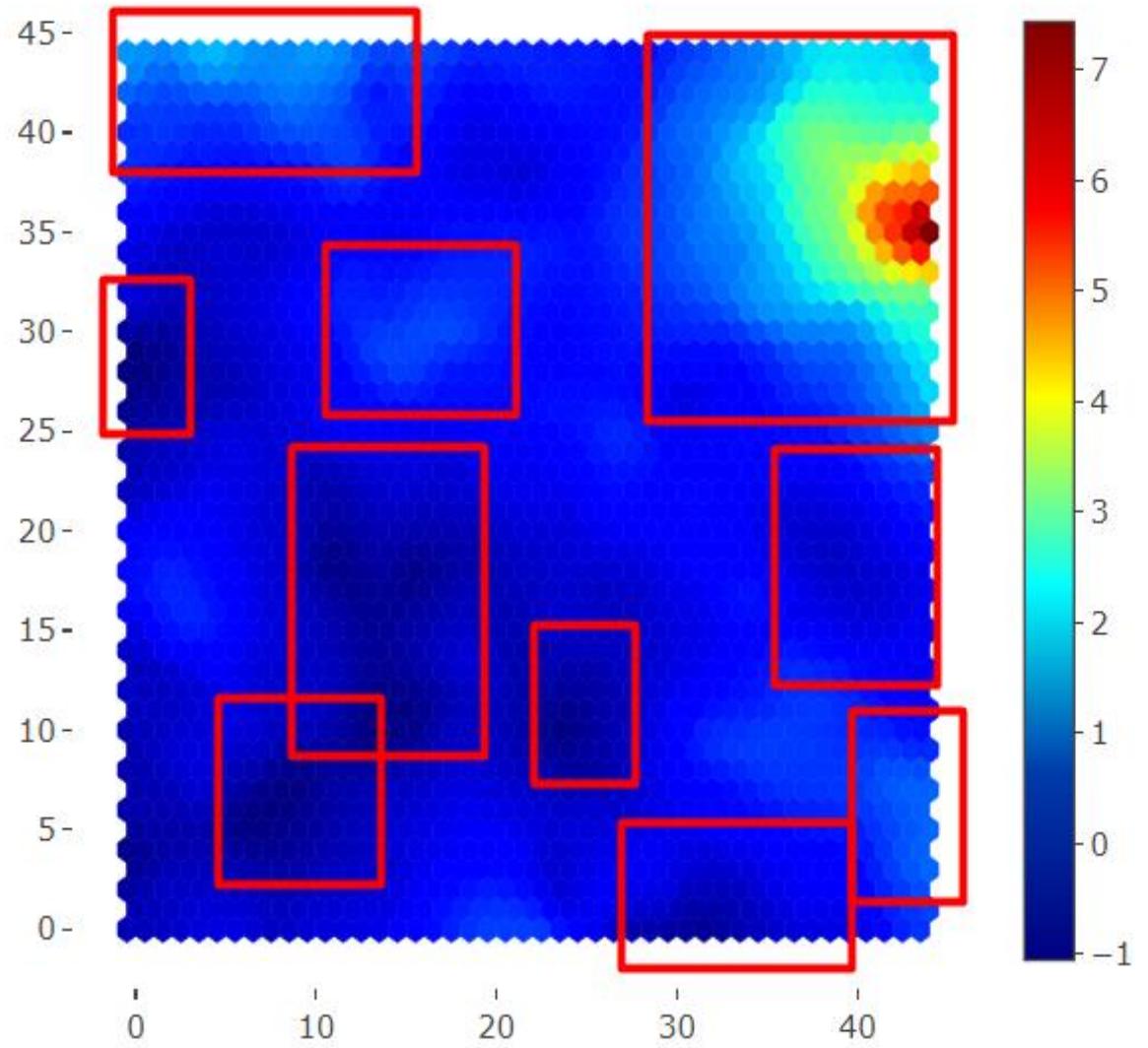


Mapa
De
Componente

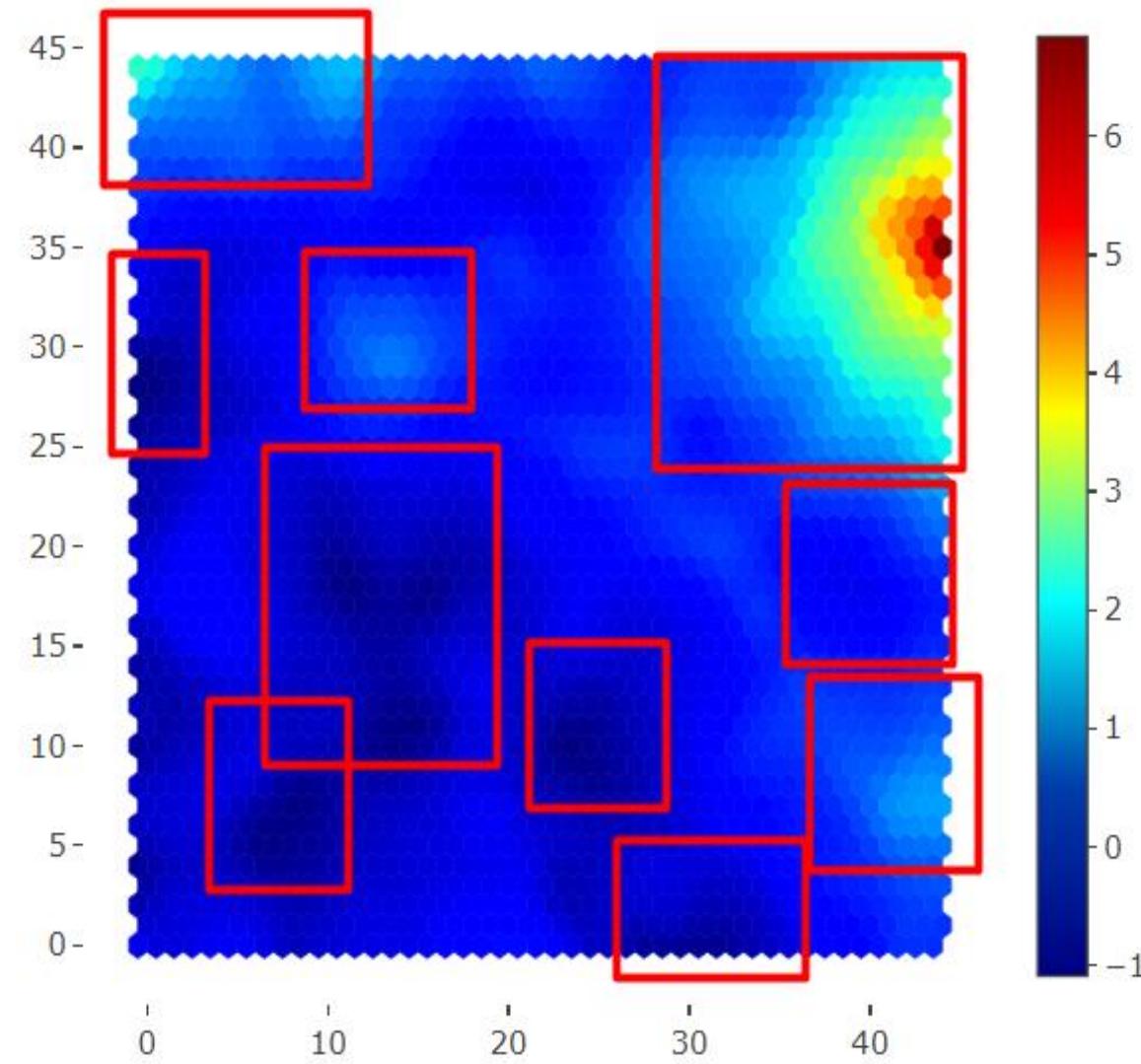


Relación Entre El Precio Medio De La Vivienda En California Y Los Ingresos Medios

total_rooms

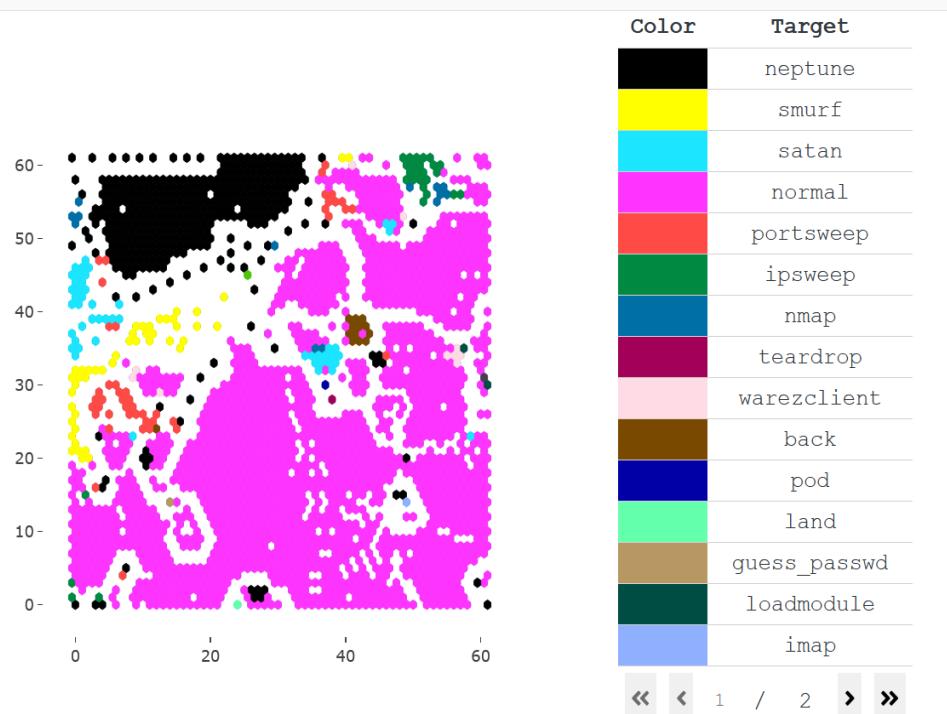


total_bedrooms

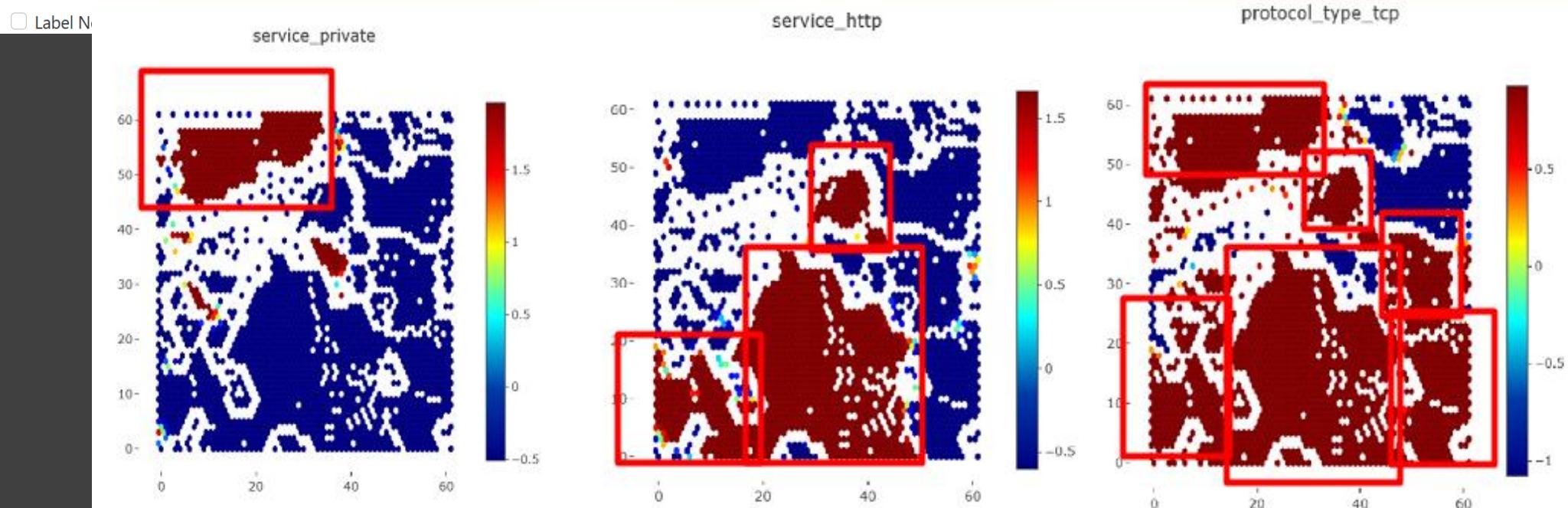


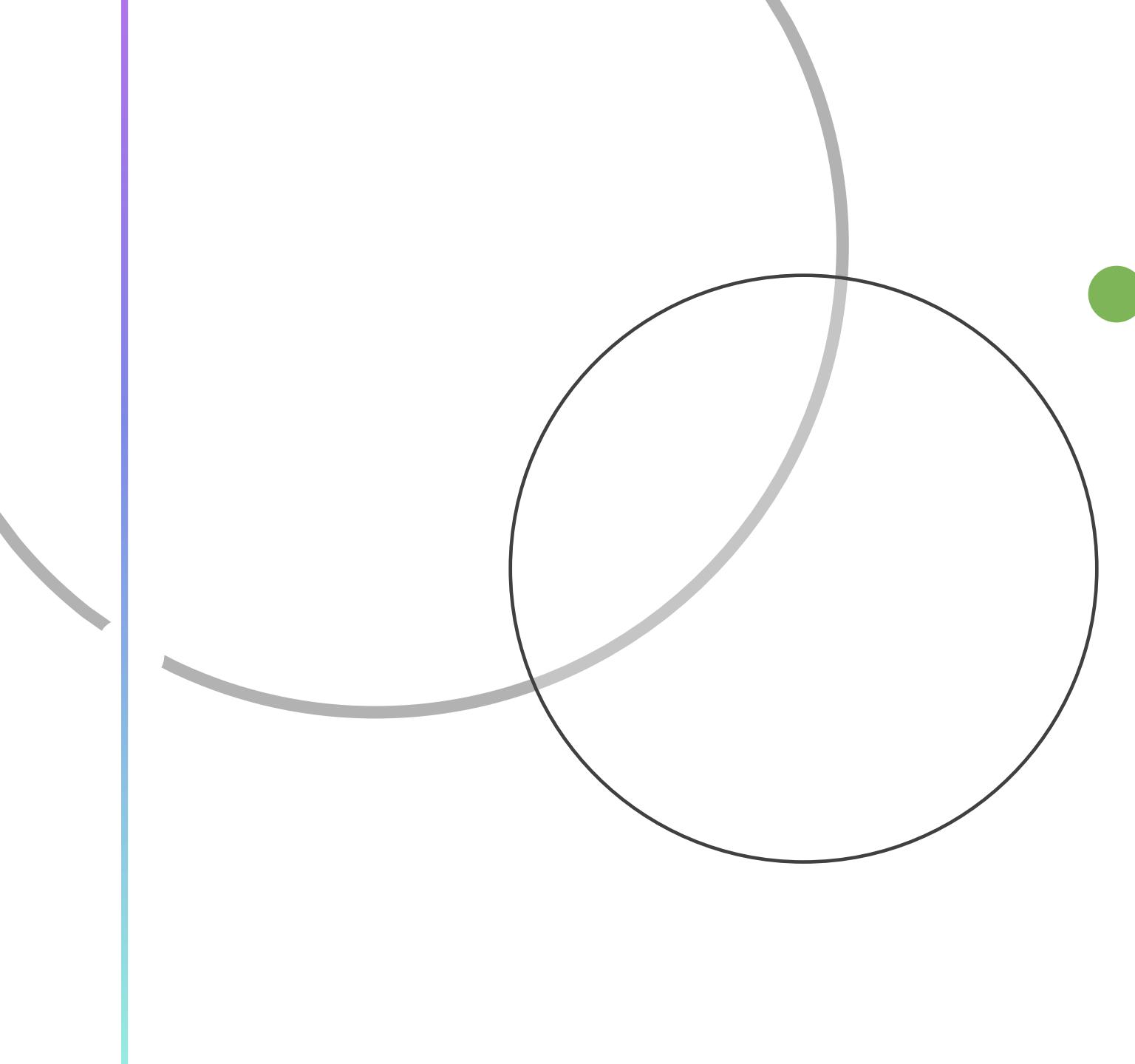
SOM Hexagonal

KDD: Ataques en Red



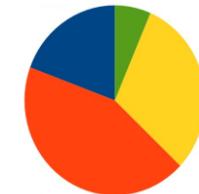
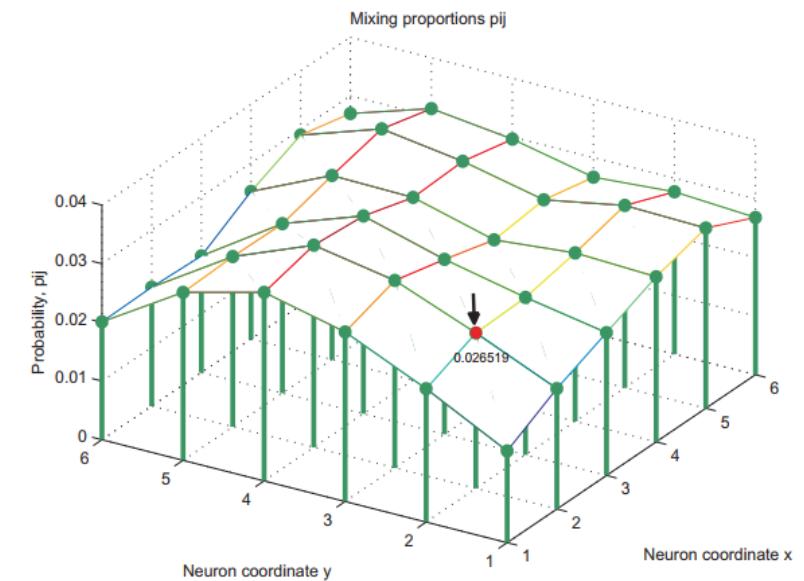
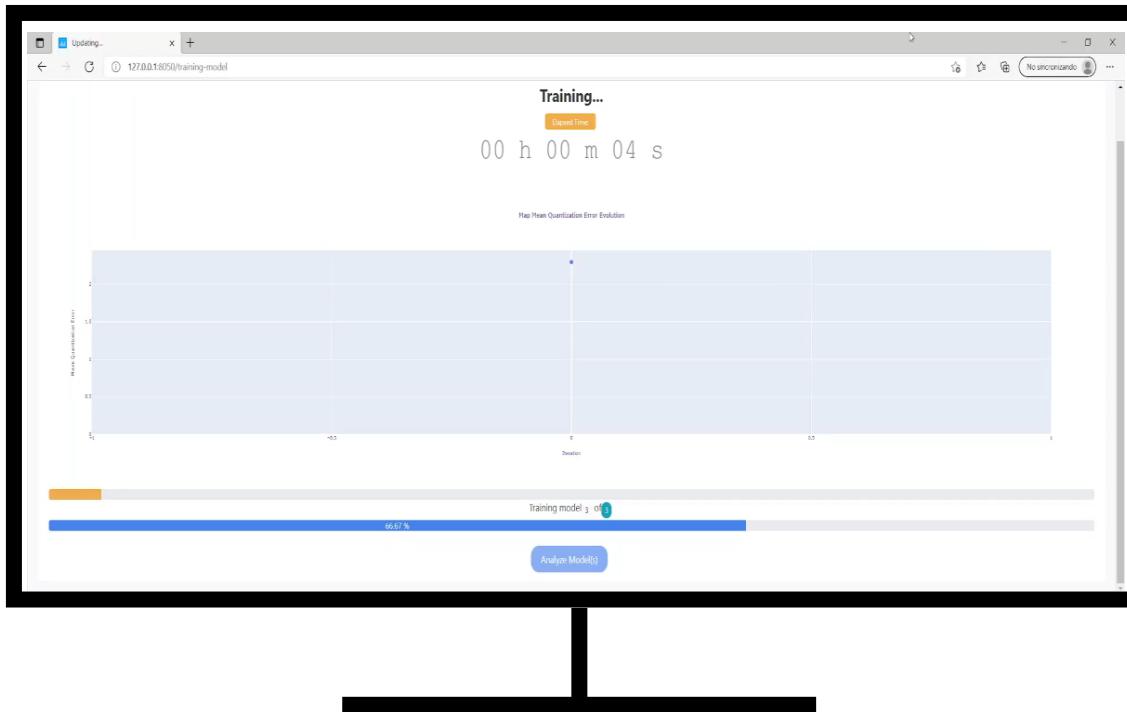
KDD Dataset: Ataques en Red





Trabajo
Futuro

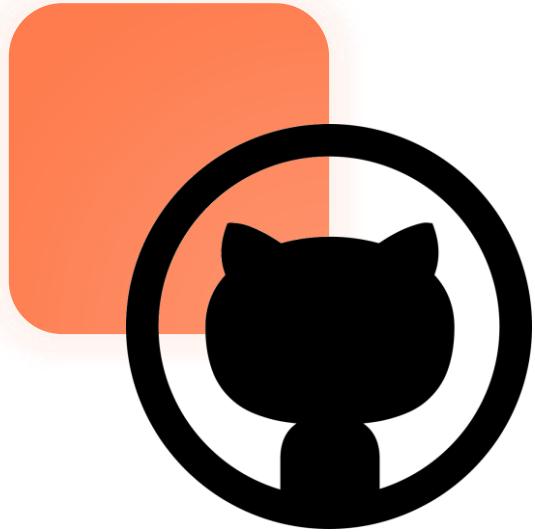
CONCLUSIONES



Referencias de Figuras

- [1] M. Riveiro, G. Falkman, and T. Ziemke, “Visual Analytics for the Detection of Anomalous Maritime Behavior,” 2008, pp. 273–279, doi: 10.1109/IV.2008.25 [Online]. Available: <https://ieeexplore.ieee.org/document/4577959>
- [2] Miljković, Dubravko. (2017). Brief Review of Self-Organizing Maps. 10.23919/MIPRO.2017.7973581.
- [3] <https://commons.wikimedia.org/wiki/File:Somtraining.svg>
- [4] <https://commons.wikimedia.org/wiki/File:TrainSOM.gif>
- [5]
- “SOM_TOOL_BOXdocumentacion_de_matlab_dice_entrenamiento_lotes.” .
- [6] https://en.wikipedia.org/wiki/File:Gaussian_2d_surface.png
- [7], [8], [9] A. Rauber, Dieter Merkl, and M. Dittenbach, “The Growing Hierarchical Self-Organizing Map: Exploratory Analysis of High-Dimensional Data,” ResearchGate, Nov. 2002. https://www.researchgate.net/publication/200744764_The_Growing_Hierarchical_SelfOrganizing_Map_Exploratory_Analysis_of_High-Dimensional_Data (accessed Jan. 06, 2021).





Repositorio

https://github.com/adriwitek/som_analytic_tool

Gracias por su atención