

Análisis Exploratorio de Datos

6A. Aplicación de la Red Neuronal SOM como método de proyección y clustering

Nombre: David Aaron Ramirez Olmeda

Programa: Maestría en Ciencia de Datos e Información



Informe sobre la Aplicación de la Red Neuronal SOM

Introducción:

En esta unidad, exploramos el uso de LabSOM, una poderosa herramienta para la visualización de datos y el agrupamiento de neuronas en redes. Para llevar a cabo este ejercicio, optamos por utilizar un conjunto de datos de jugadores de FIFA 2024 disponible en Kaggle. Elegimos este conjunto de datos debido a su riqueza en atributos numéricos, lo que lo hizo idóneo para nuestro propósito de estudio. A lo largo de este informe, describiremos el proceso que seguimos y los resultados que obtuvimos al aplicar la Red Neuronal SOM a estos datos.

Análisis y Resultados:

Iniciamos nuestra investigación reduciendo el conjunto de datos para simplificarlo y seleccionando un subconjunto de variables relevantes. Esto nos permitió trabajar con aproximadamente 600 filas de datos. A continuación, resumimos las etapas clave de nuestro análisis

Carga de Datos:

Importamos el conjunto de datos en LabSOM y procedimos a configurar la red neuronal.

LabSOM

File View Edit Tools Help

Exp1

Length: 30 Width: 20 Type: Training 500 Mix Inc: 100 - RTSOM - all - all RB

Data SOM Errors Maps Maps GL Components Maxima Components Clustering Test Data

Preview test file C:\Users\ddan\Downloads\archive\player_stats.txt

country height weight age ball_control aggression reactions interceptions vision composure crossing shot_pass long_pass acceleration stamina strength balance sprint speed agility jumping heading shot_power finishing long_shots curve_passes Senegal 189.80 30.68 61.64 22.50 64.33 57.42 53.54 36.58 68.60 75.76 74.71 58.53 72

Colombia 192.84 23.55 72.58 65.10 55.33 64.49 41.50 86.45 62.43 51.64 54.30 31.32 41

Brazil 181.80 31.89 62.70 70.64 54.60 63.64 87.81 42.67 65.65 54.60 64.68 65.48

Germany 193.84 29.27 65.20 65.48 14.35 18.46 36.56 41.48 34.65 17.51 14.25 20.26

Austria 187.86 33.46 71.64 66.29 70.44 58.53 35.73 82.56 63.67 80.67 32.34 33.23 22

England 191.80 23.54 28.12 35.34 11.23 20.38 39.64 24.31 34.67 15.48 4.63 16

Portugal 183.83 31.20 27.70 21.55 44.10 25.23 55.44 75.65 54.62 72.21 48.14 16 15 13

Switzerland 194.88 25.52 63.61 68.34 56.43 60.55 50.78 94.53 68.34 74.74 41.33 18 25 34

Portugal 185.75 20.41 46.54 56.25 46.26 49.45 63.64 78.63 55.45 71.55 39.21 24 26 39

Separator ☐ Preview

☒ First row as component names Indicators in the last 0 columns Vectors: 600

☒ First column as data labels Dimension: 25

	height	weight	age	ball_control	aggression	reactions	interceptions	vision	composure	crossing	shot_pass	long_pass	acceleration	stamina	strength	balance	sprint
Colombia	192	84	22	55	72	68	65	30	50	33	64	49	41	55	86	40	52
Brazil	181	80	31	69	62	70	70	64	54	60	63	63	64	87	81	42	67
Germany	193	84	29	25	27	65	20	49	48	14	35	18	46	38	60	41	40
Austria	187	86	33	46	71	64	66	29	70	44	58	53	35	73	82	56	63
England	191	80	23	14	28	50	12	38	34	11	23	20	38	28	64	34	31
Portugal	183	83	31	20	27	70	21	55	44	10	25	23	55	44	75	65	54
Switzerland	194	88	25	52	63	61	60	34	56	43	60	55	55	75	84	33	60
Portugal	185	75	20	41	46	54	56	25	46	26	49	45	63	64	78	63	55
Senegal	189	80	30	68	61	64	22	50	64	33	57	42	53	54	86	58	69

LabSOM VBBioSOM

Iteration: Finished training

Configuración de la Red:

Observamos el progreso computacional de la creación de la red neuronal y examinamos las neuronas generadas junto con sus atributos asociados.

The image displays the LabSOM software interface. The main window shows the SOM configuration parameters:

- Length: 30
- Width: 20
- Type: [Icon]
- Training: [Icon]
- 500
- Mix: [Icon]
- Inc: 100
- RTSOM: + all - all RB

The SOM configuration window is open, showing the following parameters:

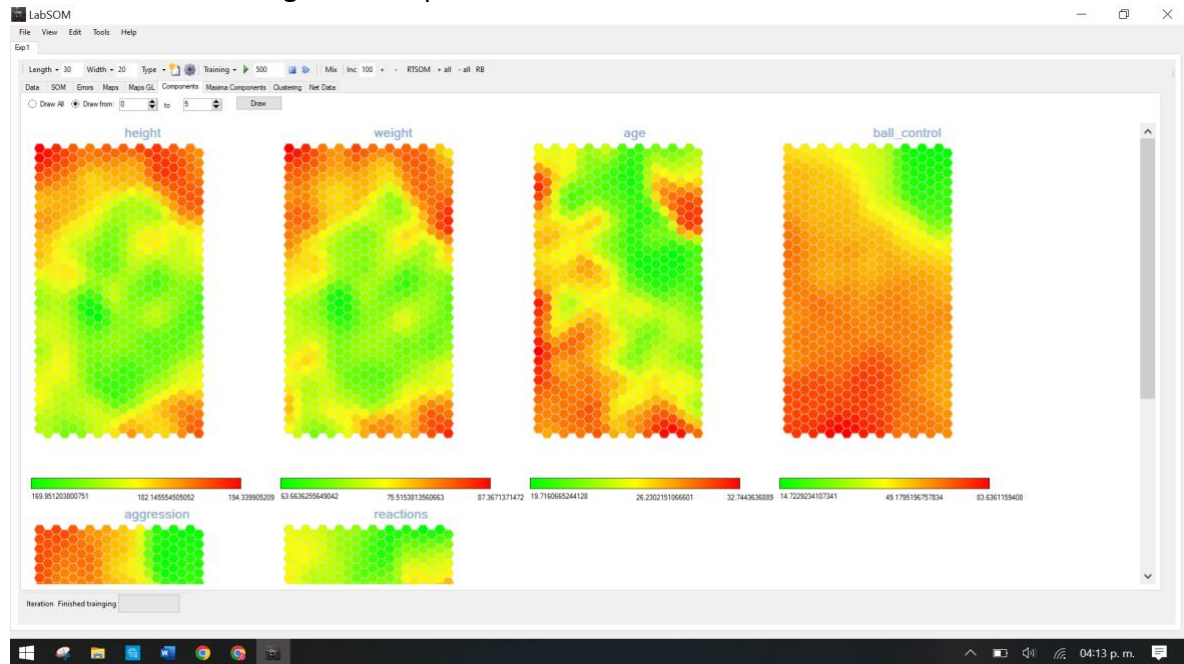
- Correction Functions: Kohonen
- Distancia: Euclidean
- Learning Factor: ☒ Continuo
- Max Value: 0.9
- Time to Change: 10
- Value to Change: 0.02
- Sigma: ☒ Continuous (From Initial Radius to 1 or 0)
- Initial Radius (2 is the half of the map): 2
- Time to Change: 10
- Part of Radius after Time to Change: 2
- ☐ Allow Sigma going to Cero
- ☐ Normalize Batch

The system resource monitor window is also open, showing the following information:

- CPU: 100% 2.96 GHz
- Memoria: 3.8/7.0 GB (54%)
- Disco 0 (C:): SSD 0%
- Wi-Fi: E: 16.0 R: 0 Kbps
- GPU 0: AMD Radeon(TM) Vega... 9% (68 °C)
- CPU AMD Ryzen 5 2500U with Radeon Vega Mobi...
- Uso: 100%
- Velocidad: 2.96 GHz
- Procesos: 200
- Subprocesos: 2701
- Identificadores: 93124
- Velocidad de base: Sockets: Núcleos: Procesadores lógicos: Virtualización: Compatibilidad con Hyper-V: Caché L1: Caché L2: Caché L3:
- Tiempo activo: 15:18:37:45

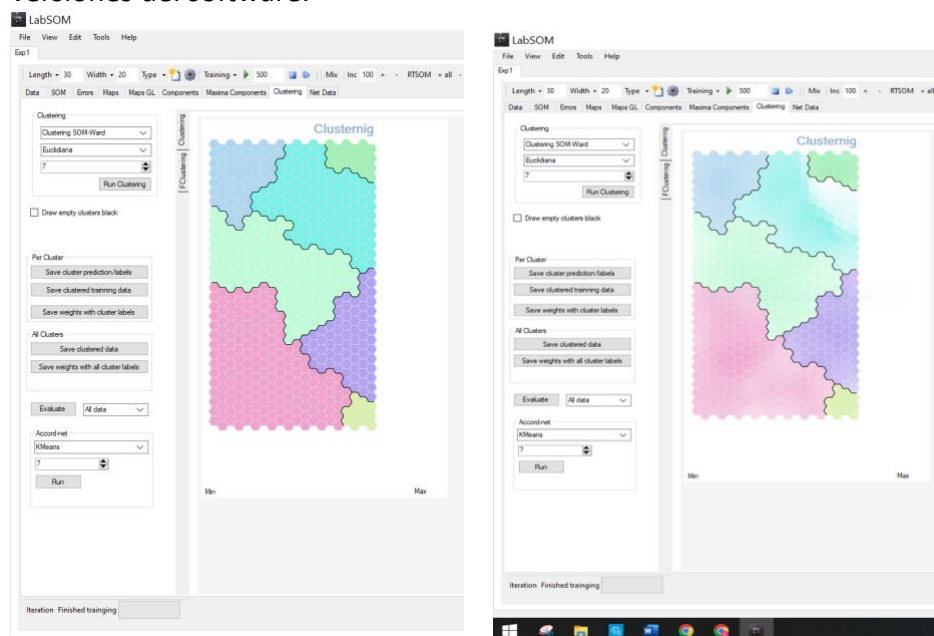
Análisis de Mapas de Componentes:

Exploramos los mapas de componentes que corresponden a cada variable significativa de nuestro conjunto de datos. Además, aplicamos transformaciones a estos mapas, como el uso de mediciones logarítmicas para resaltar o suavizar los valores.



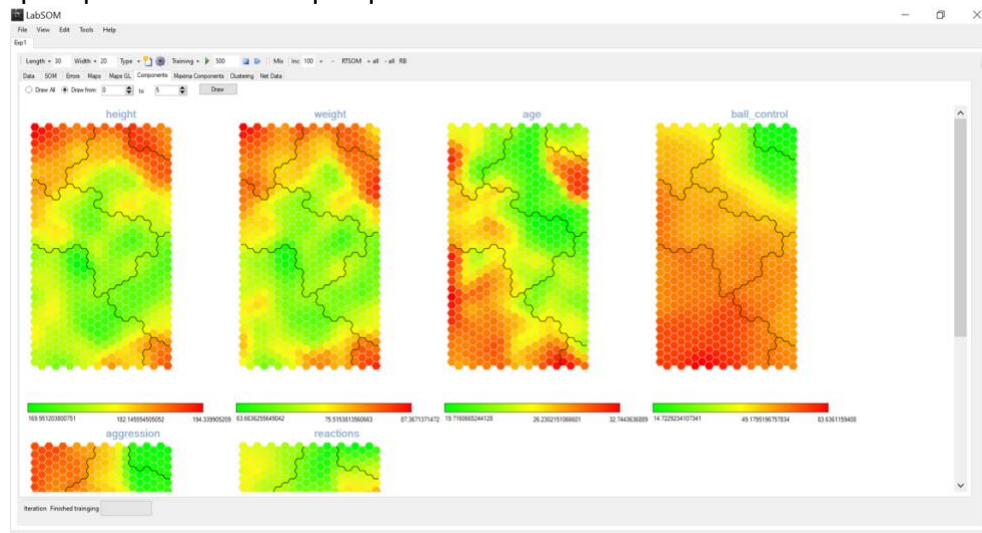
Clustering:

Luego, procedimos al clustering, donde analizamos cómo se comportan los agrupamientos al variar el número de clusters. Sin embargo, notamos que LabSOM no proporciona métodos automáticos para determinar el número óptimo de clusters, lo que nos obligó a realizar esta evaluación de manera visual. Este aspecto podría ser mejorado en futuras versiones del software.



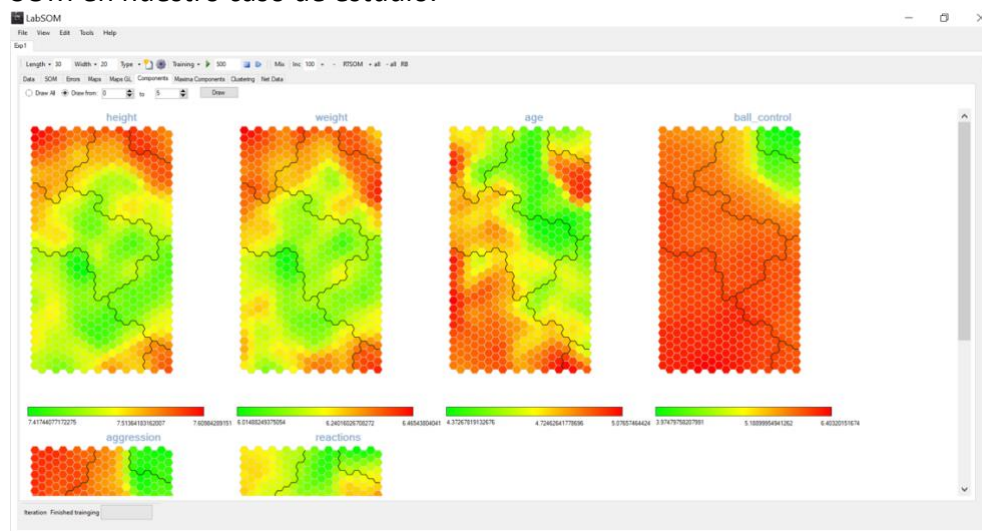
Mejora de la Visualización:

Durante el proceso de análisis, identificamos la necesidad de mejorar la visualización de nuestros resultados. Para lograr esto, realizamos modificaciones selectas en la configuración de LabSOM y en la representación gráfica de los resultados. Al realizar estas mejoras, notamos una clara diferencia en la claridad y comprensión de nuestros hallazgos. Estas modificaciones incluyeron la segmentación de grupos, la adición de etiquetas y la aplicación de escalas logarítmicas para resaltar detalles importantes en los datos. Esta experiencia nos demostró que, con unas pocas pero poderosas modificaciones, es posible optimizar significativamente la visualización de los resultados de la Red Neuronal SOM, lo que aporta una valiosa perspectiva a nuestro análisis.



Comparación de Ward SOM con Otros Métodos:

Además de utilizar Ward SOM, consideramos la posibilidad de comparar nuestros resultados con otros métodos de proyección y agrupamiento sobre los datos originales. Esto podría brindar una perspectiva más completa sobre la efectividad de la Red Neuronal SOM en nuestro caso de estudio.



Conclusiones:

En el transcurso de esta actividad, llegamos a las siguientes conclusiones:

- LabSOM se reveló como una herramienta poderosa y completa para el análisis de datos y el clustering mediante Redes Neuronales SOM. Su capacidad para visualizar y agrupar datos resultó valiosa en nuestro análisis.
- La elección de utilizar un conjunto de datos de jugadores de FIFA 2024 demostró ser adecuada, ya que sus atributos numéricos permitieron un análisis eficaz.
- Experimentamos con mejoras en la visualización de resultados, lo que destacó la importancia de la representación gráfica en el análisis de datos. Pequeñas modificaciones lograron una mejora significativa en la comprensión de los resultados.
- Esta actividad nos permitió explorar alternativas a las herramientas que utilizamos comúnmente y ampliar nuestro enfoque en el análisis de datos.
- Sin embargo, un punto de mejora que identificamos es la limitación de LabSOM para usuarios de sistemas Mac. Hubiera preferido que la herramienta fuera compatible con esta plataforma, lo que ampliaría su accesibilidad y utilidad para una gama más amplia de usuarios.

Referencias:

En esta actividad, utilizamos LabSOM como la principal herramienta de análisis de datos. Puedes encontrar más información sobre LabSOM en
[<https://www.dynamics.unam.edu/DinamicaNoLineal3/labsom.html>]