EDUCACIÓN TECNOLÓGICO NACIONAL DE MÉXICO
IPSUMTEC
ISSN: 2594 - 2905

APLICACIÓN DE LA PSEUDODERIVADA PARA LA IDENTIFICACIÓN DE PUNTOS DE INFLEXIÓN EN CONJUNTOS DE DATOS

APPLICATION OF PSEUDODERIVATIVE FOR THE IDENTIFICATION OF INFLECTION POINTS IN DATASETS

López Nava Misael[1], Reyes Reyes Juan[2], Osorio Gordillo Gloria Lilia[3], Astorga Zaragoza Carlos Manuel[4], Muñiz Rascado Luis José[5]

https://doi.org/10.61117/ipsumtec.v8i2.386

[1]Tecnológico Nacional de México/CENIDET, d21ce044@cenidet.tecnm.mx, https://orcid.org/0009-0005-0736-280X.
[2]Tecnológico Nacional de México/CENIDET, juan.rr@cenidet.tecnm.mx, https://orcid.org/0000-0001-7415-9987.
[3]Tecnológico Nacional de México/CENIDET, gloria.og@cenidet.tecnm.mx, https://orcid.org/0000-0002-3445-8867.
[4]Tecnológico Nacional de México/CENIDET, carlos.az@cenidet.tecnm.mx, https://orcid.org/0000-0002-4678-9006.
[5]Tecnológico Nacional de México/Instituto Tecnológico de Zacatepec, luis.mr@zacatepec.tecnm.mx, https://orcid.org/0000-0003-2941-2482.

Resumen -- En el ámbito del reconocimiento de patrones existen técnicas para identificar las características deseadas de un conjunto de datos, en otras palabras, clasificar la información en categorías predefinidas. La pseudoderivada es una herramienta que puede ser utilizada en el reconocimiento de imágenes, reconocimiento de voz, reconocimiento de escritura a mano, etc. Este trabajo presenta la aplicación de la pseudoderivada para la identificación de características en conjuntos de datos, basada en un algoritmo para el tratamiento de datos fundamentado en ecuaciones diferenciales lineales, mismo que permite procesar un dato en un momento específico a partir del análisis de n datos inmediatos anteriores y posteriores, de tal manera que se puede ampliar o disminuir el número de elementos para su manipulación. La pseudoderivada no solo permite identificar mínimos y máximos en una primera aplicación, sino que, además de esta información, al aplicarse una segunda vez al mismo conjunto de datos, permite obtener los puntos de inflexión de la información sobre la que opera.

Palabras Clave: Grandes Datos, Minería de Datos, Reconocimiento de patrones.

Abstract -- In the field of pattern recognition, there are techniques to identify the desired characteristics of a data set, in other words, to classify the information into predefined categories. The pseudoderivative is a tool that can be used in image recognition, voice recognition, handwriting recognition, etc. This work presents the application of the pseudoderivative for the identification of characteristics in data sets, based on an algorithm for data processing based on linear differential equations, which allows processing a data at a specific time from the analysis of n immediately preceding and following data, in such a way that the number of elements for manipulation can be increased or decreased. The pseudoderivative not only allows identifying minimums and maximums in a first application, but, in addition to this information, when applied a second time to the same data set, it allows obtaining the inflection points of the information on which it operates.

Key words – Big Data, Data Mining, Pattern Recognition.

INTRODUCCIÓN

El concepto de Big Data se refiere a “datos masivos” y a menudo no estructurados, en los que las capacidades de procesamiento de las herramientas tradicionales de gestión de datos resultan ser inadecuadas. Big Data puede ocupar terabytes y petabytes de espacio de almacenamiento en diversos formatos, incluidos texto, video, sonido, imágenes y más. Aunque el término Big Data es relativamente nuevo, la tendencia a agrupar y almacenar grandes volúmenes de información para análisis a futuro es muy antigua. [1] [2]

Big data es un conjunto de datos de una gran variedad y formatos; que se acumulan en grandes volúmenes y a una velocidad cada vez mayor. A esto, es lo que se conoce como las 3 V’s (dimensiones) de la Big Data. Estas dimensiones se describen a continuación:

Volumen. Las organizaciones recopilan datos de una amplia variedad de fuentes, incluidas transacciones financieras, redes sociales, sensores o máquina a máquina. En el pasado, el almacenamiento hubiera sido un problema, pero las nuevas tecnologías facilitan la tarea.

Velocidad. Los datos fluyen a una velocidad sin precedentes y, por lo tanto, deben gestionarse de manera oportuna. El uso cada vez más frecuente de etiquetas RFID (identificaciones de radiofrecuencia), sensores y medición inteligente (sistemas de lectura de contadores) aumentan la necesidad de gestionar flujos de datos en tiempo real o casi.

Variedad. Los datos llegan en cualquier formato, desde datos estructurados y numéricos en bases de datos tradicionales a no estructuradas, como documentos de

texto, correo electrónico, video, audio, datos de cotizaciones y transacciones financieras. [3][4]

Sin embargo, con la generación de información tan vertiginosa, se han considerado otras tres dimensiones: variabilidad, que se refiere al aumento exponencial de la velocidad y variedad de datos se combina con el hecho de que los flujos pueden ser muy inconsistentes y con picos periódicos; veracidad, derivado a que es de suma importancia determinar si el dato analizado es relevante o no para el análisis de la información; y valor, por la capacidad de los datos para generar información útil y estratégica para las organizaciones, que les permita tomar mejores decisiones, mejorar la eficiencia y lograr un mayor valor empresarial. [5] [6]

La minería de datos estudia métodos y algoritmos que permiten la extracción automática de información sintetizada que permite caracterizar las relaciones escondidas en la gran cantidad de datos; también pretende que la información obtenida posea capacidad predictiva, facilitando el análisis de los datos de forma eficiente.[7]

Bajo la denominación de "minería de datos" se han agrupado diversas técnicas estadísticas y del aprendizaje automático enfocadas, principalmente, a la visualización, análisis, y modelización de información de bases de datos masivas. [8] [9]

Se propone un método para la extracción de características de imágenes basado en pseudoderivadas. Las pseudoderivadas son una generalización de las derivadas tradicionales que pueden capturar información más rica sobre la textura local de una imagen. El método propuesto utiliza pseudoderivadas de diferentes órdenes para calcular un conjunto de características que luego se utiliza para la clasificación de imágenes. El método se basa en dos etapas principales: 1) cálculo de pseudoderivadas: Se calculan pseudoderivadas de diferentes órdenes para cada píxel de la imagen. 2) extracción de características: Se utilizan las pseudoderivadas para calcular un conjunto de características que describen la textura local de la imagen. El método propuesto se evaluó en una serie de conjuntos de datos de imágenes y se comparó con otros métodos de extracción de características. Los resultados mostraron que el método propuesto logra una mejor precisión de clasificación en la mayoría de los casos. El método de extracción de características basado en pseudoderivadas es una herramienta prometedora para la clasificación de imágenes. [10]

El método es capaz de capturar información más rica sobre la textura local de una imagen que las derivadas tradicionales, lo que conduce a una mejor precisión de clasificación. El método propuesto es computacionalmente más costoso que otros métodos de extracción de características. Además, el rendimiento del método puede verse afectado por el ruido en las imágenes. El método propuesto se puede utilizar en una variedad de aplicaciones de clasificación de imágenes, como el reconocimiento de objetos, la clasificación de escenas y la detección de defectos.

Se propone otro método para la identificación de hablantes utilizando pseudoderivadas, que son características extraídas de la señal de voz que son robustas a las variaciones del habla. Los autores primero extraen pseudoderivadas de la señal de voz. Luego, utilizan estos pseudoderivadas para entrenar un modelo de identificación de hablantes. Finalmente, evalúan el rendimiento del modelo en una base de datos de habla. Los resultados muestran que el método propuesto logra una precisión de identificación de hablantes comparable a los métodos de última generación. Además, el método propuesto es más robusto a las variaciones del habla, como el ruido y el cambio de canal. [11]

El método propuesto es una alternativa prometedora a los métodos tradicionales de identificación de hablantes. Su robustez a las variaciones del habla lo hace adecuado para una variedad de aplicaciones, como la autenticación de voz y el reconocimiento de voz. El estudio se basa en una sola base de datos de habla. Se necesitan más investigaciones para evaluar el rendimiento del método propuesto en otras bases de datos de habla. Por esta razón esta metodología se puede utilizar en una variedad de aplicaciones, como la autenticación de voz, el reconocimiento de voz y la síntesis de voz.

Se propone otro método para el reconocimiento de dígitos manuscritos utilizando pseudoderivadas, las cuales son características locales que se pueden calcular de forma rápida y eficiente, y que son robustas a la distorsión y al ruido. El método propuesto se evaluó en una base de datos de imágenes de dígitos manuscritos y se comparó con otros métodos existentes. Los resultados mostraron que el método propuesto tiene una precisión comparable a los métodos existentes, pero con una complejidad computacional menor. [12]

Además, esta metodología es una alternativa viable para el reconocimiento de dígitos manuscritos. Es rápido, eficiente y robusto a la distorsión y al ruido, es novedosa y tiene un buen potencial para mejorar la precisión del reconocimiento de dígitos manuscritos. Los resultados experimentales son convincentes y muestran que este método es comparable a los métodos existentes. El artículo tiene algunas limitaciones, como la falta de pruebas en otras bases de datos de imágenes de dígitos manuscritos. [13]

En lo que se refiere a un punto de inflexión, este alude a dos componentes: por una parte, un cambio en la dirección (inflexión) y un momento (punto), es decir, ocurren en un tiempo específico (cronológico). [14]

Se pueden identificar de acuerdo a su posicionamiento, es decir, concavidad hacia abajo o concavidad hacia arriba, lo cual describe un comportamiento cuando un fenómeno a partir de un momento preciso tiende a decrecer (concavidad hacia abajo), o bien, cuando un este fenómeno tiende a crecer (concavidad hacia arriba). Estos comportamientos se obtienen del análisis de los datos subsecuentes y anteriores a un punto específico en el tiempo. En esta interpretación recae la importancia del análisis de información en el presente trabajo, pues se propone su aplicación en conjuntos de datos reales, para

Recepción: 12/06/2025 | Aceptación: 13/09/2025 | Publicación: 08/10/2025

ISSN: 2594 - 2905

este caso las muertes por COVID-19 del 2020 al 2023. [15]

DESARROLLO

Metodología.

La función sub es una función previa a la segmentación de toda una variable colocada en un valor k y a través de esta, se puede obtener información de un k-ésimo dato, mediante el promedio de los vecinos previos y subsecuentes. Esta función está representada por la ecuación 1.

$$sub_p(v,k) \coloneqq [v\left(k-\frac{p-1}{2}\right),\dots,v(k-1),v(k),v(k+1),\dots,v\left(k+\frac{p-1}{2}\right)] \quad (1)$$

Donde:

v es cada uno de los valores a utilizar para el procesamiento.

k es el instante de tiempo del dato.

Con esta representación, se puede analizar un número impar de datos, en proporción al número de elementos que componen el conjunto de datos. En el caso de una ventana de 5 datos, se puede representar mediante la ecuación 2.

$$sub_5(v,k) \coloneqq [v(k-2),v(k-1),v(k),v(k+1),v(k+2)] \quad (2)$$

En la etapa de segmentación se realizan procesamientos para extraer características, para el presente trabajo se utilizó el filtro pseudoderivada.

El filtro pseudoderivada se puede representar mediante un vector que como el de la ecuación 3, y al aplicarlo sobre un conjunto de datos, permite identificar: cuando hay un cruce por cero, se identifica un máximo o un mínimo, y cuando permanece cerca de cero, es constante.

$$z(k) = sub_5(b_1^l,k)\begin{bmatrix}-1\\-1\\0\\1\\1\end{bmatrix}, k = 3,4,\dots,t_f-2 \quad (3)$$

Donde:

$z(k)$ es la primer pseudoderivada.

b_1^l es el conjunto de datos sobre el que se va aplicar el filtro.

k es el instante de tiempo del dato.

t_f-2 es el último valor a considerar en el tratamiento.

Esta función va a permitir identificar valores máximos y mínimos, cuando exista un cruce por cero, observándose en la gráfica de los resultados del procesamiento.

En Matemáticas, un punto de inflexión de una función, es un punto donde los valores de una función continua en x pasan de un tipo de concavidad a otra. La curva atraviesa la tangente. La segunda derivada de la función f en el punto de inflexión es cero, o no existe. Con base en lo anterior, se puede aplicar una segunda ocasión la función pseudoderivada sobre el mismo conjunto de datos para obtener los puntos de inflexión del mismo. En este sentido, la segunda opera sobre el vector resultante de la primera aplicación sobre el conjunto de datos original.

De tal forma, la representación de la segunda pseudoderivada sería como en la ecuación 4.

$$z'(k) = sub_5(sub_5(b_1^l,k))\begin{bmatrix}-1\\-1\\-1\\-1\\0\\1\\1\\1\\1\end{bmatrix}, k = 5,6,\dots,t_f-4 \quad (4)$$

Donde:

$z'(k)$ es la segunda pseudoderivada.

b_1^l es el conjunto de datos sobre el que se va aplicar el filtro.

k es el instante de tiempo del dato.

t_f-4 es el último valor a considerar en el tratamiento.

Una vez que se obtiene el vector de características con los puntos de inflexión identificados en el conjunto de datos, se propone una función que permita identificar de manera automática los puntos de inflexión. Para esta consideración es importante determinar el tamaño adecuado de la máscara, pues el tamaño de ésta, genera una aproximación más fidedigna del patrón (punto de inflexión) en el conjunto de datos analizado.

Un aspecto fundamental para la formulación de una función automatizada que permita el tratamiento de los datos sin necesidad de obtener una versión visual (gráfica) y que determine un parámetro para identificar un punto de inflexión, independientemente del número de muestras del conjunto de datos. Este valor se va a identificar como ε, y que se obtendrá para poder discriminar los valores obtenidos de la segunda pseudoderivada, con la finalidad de dar una característica de automatización.

La función para identificar que en ese punto del conjunto de datos existe un punto de inflexión, es la que se muestra como ecuación 5.

$$g(k) = \begin{cases}1, |z'(k)| < \varepsilon\\ 0, |z'(k)| \geq \varepsilon\end{cases} \quad (5)$$

Donde:

$g(k)$ es la función para identificar los puntos de inflexión.

ε es el valor para discriminar los valores de z´(k).

Con esta función lo que se pretende es obtener un valor 1 para identificar un punto de inflexión y 0 cuando no lo haya.

A continuación, se proponen las funciones para discriminar los valores, con diferentes grados de rigidez para la detección de los puntos de inflexión, mediante funciones.

Criterio 1. Un décimo de la diferencia entre el valor máximo y el mínimo. Se muestra en la ecuación 6.

$$\varepsilon \coloneqq \alpha\,\frac{\max_k\{x(k)\} - \min_k\{x(k)\}}{10}, 0 < \alpha \leq 1 \quad (6)$$

Este criterio identifica los puntos de inflexión sin ser tan estricto cuando el valor de α es cercano a 1, es decir, tiende a identificarlos sin necesidad de que los valores analizados sean tan cercanos a cero.

Recepción: 12/06/2025 | Aceptación: 13/09/2025 | Publicación: 08/10/2025

ISSN: 2594 - 2905

Criterio 2. Un décimo de la diferencia máxima entre dos valores contiguos. Se muestra en la ecuación 7.

$$\varepsilon := \alpha \frac{\max_k\{|x(k) - x(k+1)|\}}{10}, 0 < \alpha \leq 1 \qquad (7)$$

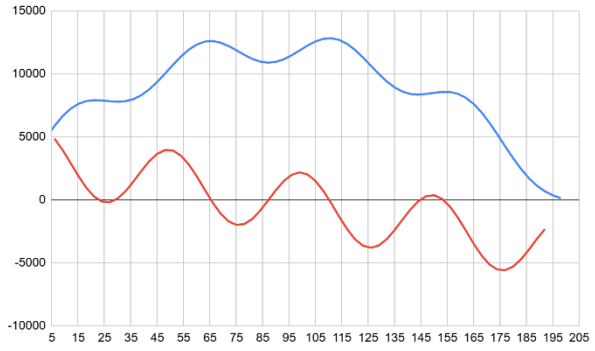
Este criterio es más estricto que el anterior.
Criterio 3. Un centésimo de la máxima magnitud absoluta de un valor. Se muestra en la ecuación 8.

$$\varepsilon := \alpha \frac{\max_k\{|x(k)|\}}{100}, 0 < \alpha \leq 1 \qquad (8)$$

Este criterio es el más estricto para identificar los puntos de inflexión.
Una vez que se aplica la función g(k) a cada uno de los valores obtenidos de la función z´(k), se genera un vector de valores 0 y 1, donde el valor 0 indica que no se trata de un punto de inflexión, mientras que un 1 indica que si lo hay.

Aplicación.
Primer conjunto de datos de prueba.
Considerando un conjunto de datos propuesto, que consta de 67 datos, tomados en el mismo número de unidades de tiempo, solo que las unidades de tiempo se incrementan de tres en tres (0 a 198), se pueden obtener los valores máximos y mínimos al aplicar la pseudoderivada al conjunto de datos.
En la siguiente figura se observa el gráfico con la línea azul que representa los valores del conjunto de datos y en color rojo el comportamiento de la pseudoderivada, sobre el conjunto de datos, se destaca que el eje x corresponde a las unidades de tiempo en que fueron tomadas las muestras.

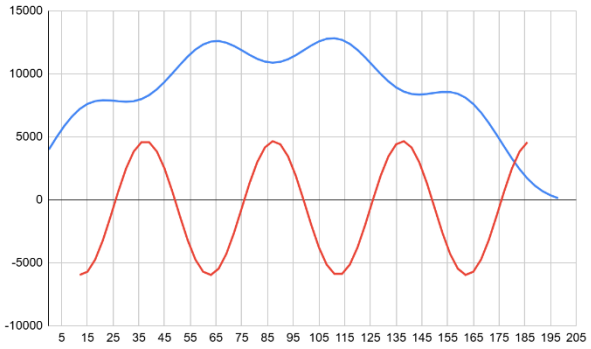
Figura 1. *Gráfica de la primer pseudoderivada*.



Fuente. *Elaboración propia*.

En la Figura 1, se pueden observar los valores máximos y mínimos, que corresponden a los cruces por cero de la línea (roja) que representa a la pseudoderivada, los valores máximos se identifican en los valores de tiempo 21, 66, 111 y 153. Mientras que, los valores mínimos se ubican en los tiempos 30, 87 y 144.
Al aplicar la segunda pseudoderivada al mismo conjunto de datos, es entonces que el comportamiento de la línea que la representa (línea roja), ahora va a identificar los puntos de inflexión en cada cruce por cero. Lo cual se observa en la Figura 2.

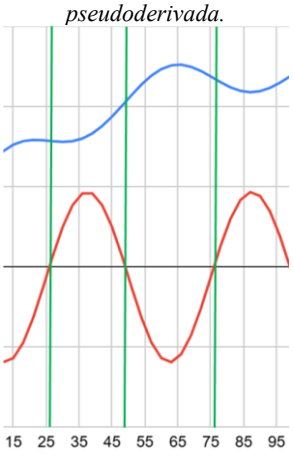
Figura 2. *Gráfica de la segunda pseudoderivada.*



Fuente. *Elaboración propia*.

En la figura 2 se puede observar que los puntos de inflexión de los datos analizados, se ubican en los valores de tiempo 27, 51, 78, 99, 126, 150 y 177. Esto se puede verificar en el fragmento tomado de la Figura 2 y que se puede observar en la Figura 3.

Figura 3. *Punto de inflexión identificado mediante la segunda pseudoderivada.*

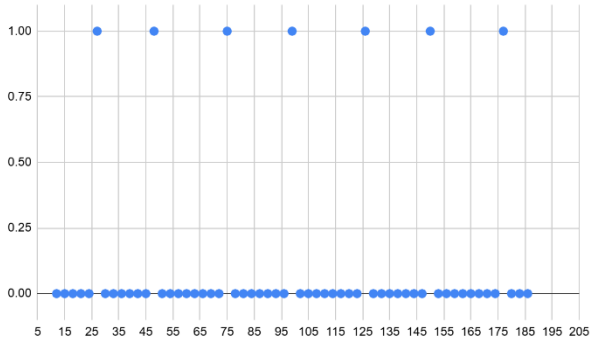


Fuente. *Elaboración propia*.

En esta gráfica, mediante líneas color verde se puede corroborar que efectivamente, a cada cruce de la pseudoderivada (línea roja) por el eje x, existe un punto de inflexión en el conjunto de datos real (línea azul), es decir, ocurre un cambio en la concavidad de la línea que representa el comportamiento de los datos analizados. En la figura 3 se identifica un cambio de concavidad negativa por positiva justo en la primera línea verde, lo contrario (cambio de concavidad hacia arriba por una concavidad hacia abajo), es decir, de positiva a negativa en la segunda línea verde. Un fenómeno similar al primero sucede en la tercera línea verde, en donde se cambia de concavidad negativa por positiva.
Esta identificación se hizo a partir de la generación de funciones propias de Google Sheets, sin embargo, es importante destacar la importancia de proponer un modelo para la identificación automática de los mismos.
Para este efecto se utilizará la función identificada como ecuación 5, con lo que se discriminarán los puntos de inflexión y posteriormente se aplicarán los criterios identificados como ecuaciones 6, 7 y 8. Esto se observa en la Figura 4.

Recepción: 12/06/2025 | Aceptación: 13/09/2025 | Publicación: 08/10/2025

Figura 4. *Puntos de inflexión identificados.*



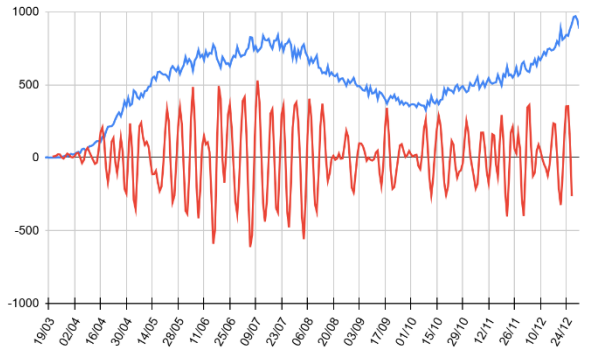
Fuente. *Elaboración propia*.

Segundo conjunto de datos de prueba.

Se realizó otro análisis de otro conjunto de datos, el cual consta de 290 muestras ordenadas por el momento en que fueron obtenidas, correspondientes a las muertes diarias por COVID-19, entre el 17 de marzo y el 31 de diciembre de 2020.

La gráfica obtenida para identificar el comportamiento del conjunto de datos y de su segunda pseudoderivada, se muestra en la Figura 5.

Figura 5. *Puntos de inflexión de las defunciones por COVID-19 durante el 2020.*



Fuente. *Elaboración propia*.

Se identifican los puntos de inflexión en diferentes fechas de la serie de datos, visualmente se identifican en cada cruce por cero de la línea roja. La forma de identificación automática, mediante las ecuaciones 5 a la 8, se muestran más adelante. La línea azul representa las defunciones diarias.

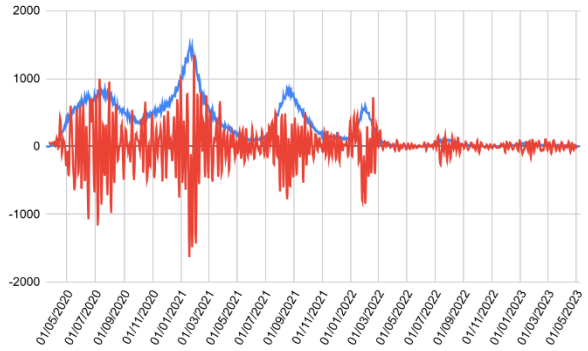
La desventaja de este tipo de representación gráfica de los puntos de inflexión es que, a mayor número de datos, es menor la identificación de los mismos, se pierde la precisión y su interpretación es complicada.

Tercer conjunto de datos de prueba.

Se utilizó un tercer conjunto de pruebas, correspondiente a las defunciones diarias por COVID-19, desde el 17 de marzo de 2020 hasta el 9 de mayo de 2023, con un total de 1,149 datos, cada uno de ellos siendo una muestra diaria.

Al obtener la gráfica de la segunda pseudoderivada y contrastarla con el comportamiento del conjunto original, se observa que, debido a la cantidad de información que se procesa, no se identifican a simple vista los puntos de inflexión, lo cual se puede observar en la Figura 6.

Figura 6. *Interfaz para la identificación de puntos de inflexión de las defunciones por COVID-19.*



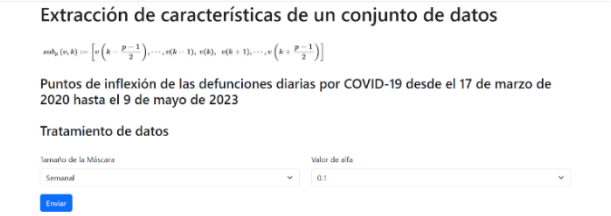
Fuente. *Elaboración propia*.

En este sentido, la identificación visual de los puntos de inflexión es muy complicada, derivado del comportamiento de los datos que se están procesando, los cuales demuestran una gran variabilidad en alternancias a la baja y al alza. Dada la cantidad de datos procesados la representación de los puntos de inflexión identificados resulta inconveniente, destacando que el modelo propuesto si los identifica.

Para este caso, se cargó la información en una hoja de Google Sheets, posteriormente se desarrolló el programa correspondiente en JavaScript, a partir del algoritmo diseñado para automatizar las funciones descritas, utilizando Google Apps Script, en el cual, también se desarrolló una interfaz en HTML.

En este sentido, esta aplicación, funciona a través de un navegador, en la que se define el tamaño de la máscara delimitada por las opciones: semanal, quincenal, mensual, trimestral, semestral y anual; y el valor de α (valores decimales de 0 a 1, con intervalos de 0.1), con estos datos se generan los valores obtenidos de la pseudoderivada y la segunda pseudoderivada, para posteriormente generar su gráfica correspondiente.

Figura 7. *Interfaz para la identificación de puntos de inflexión de las defunciones por COVID-19.*

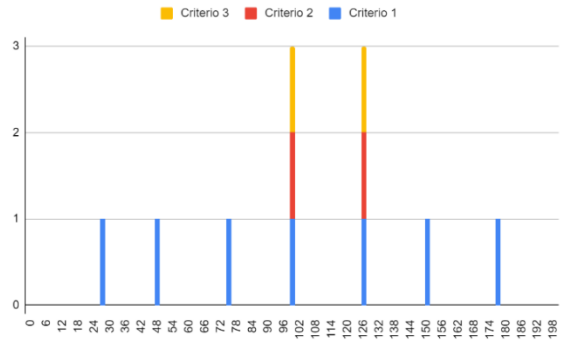


Fuente. *Elaboración propia*.

DISCUSIÓN Y ANÁLISIS DE RESULTADOS

Después de aplicar la ecuación 5 para identificar los puntos de inflexión y de los criterios definidos como las ecuaciones 6, 7 y 8, se obtuvieron las gráficas que a continuación se describen. Del primer conjunto de datos (67 valores), se obtuvo una gráfica que se muestra en la Figura 8.

Figura 8. *Puntos de inflexión identificados mediante los tres criterios utilizados en el conjunto de datos 1.*

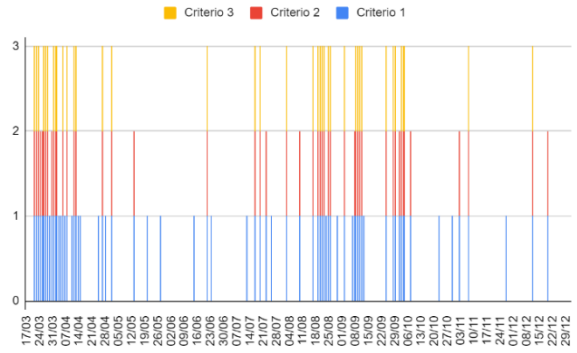


Fuente. *Elaboración propia*.

Se puede observar que mediante el criterio 1 (azul) se identificaron siete puntos de inflexión, a través del criterio 2 (rojo), solo dos, al igual que con el criterio 3 (amarillo). En la gráfica todos estos puntos están ubicados en los tiempos rotulados como eje x.

Del segundo conjunto de datos (290 valores), se obtuvo una gráfica que se muestra en la Figura 9.

Figura 9. *Puntos de inflexión identificados mediante los tres criterios utilizados en el conjunto de datos 2 (defunciones durante el 2020).*



Fuente. *Elaboración propia*.

Una vez que se aplican cada uno de los tres criterios a los valores del conjunto del segundo conjunto de datos, se puede observar que el número de puntos de inflexión se incrementa respecto al primero, lo cual es directamente proporcional al número de muestras analizadas.

Se identificaron 69 puntos de inflexión con el criterio 1 (azul), 74 puntos con el criterio 2 (rojo) y 2 con el criterio 3 (amarillo), en un análisis semanal, es decir con una tamaño de máscara (función *sub)* igual a 7, con un valor de α=0.5 (ni estricto, ni relajado). Todos estos puntos se ubican en las fechas marcadas en el eje x. Sin embargo, este tipo de gráfica complica la identificación de los puntos de inflexión.

En la siguiente tabla se muestran los puntos de inflexión identificados en el segundo conjunto de datos, con tres diferentes los tamaños de máscara 7, 31 y 93 (semanal, mensual y trimestral, respectivamente), y con tres diferentes valores de α 0.1, 0.5 y 1 (estricto, mesurado y relajado, respectivamente), esto permite tener un panorama del comportamiento del conjunto de datos.

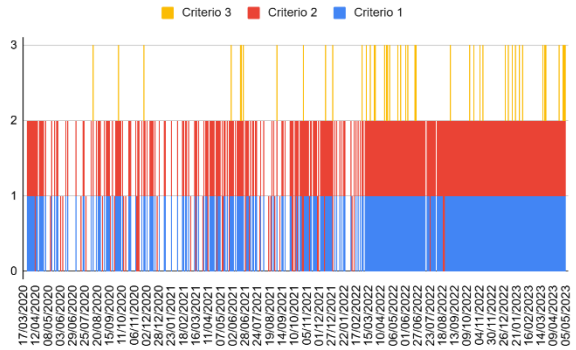
Tabla 1. *Puntos de inflexión identificados en el segundo conjunto de datos con diferentes tamaños de máscara y valores de α.*

Tamaño de la máscara	Criterio de ε	α = 0.1 (Estricto)	α = 0.5 (Mesurado)	α = 1 (Relajado)
7 (semanal)	1	15	69	120
	2	17	74	126
	3	0	2	7
31 (mensual)	1	2	15	46
	2	2	17	49
	3	0	1	1
93 (trimestral)	1	1	5	10
	2	1	5	11
	3	0	1	1

Fuente. *Elaboración propia*.

Con respecto al tercer conjunto de datos (1,149 valores) con el que se trabajó la propuesta antes mencionada, dependiendo del tamaño de la máscara y del valor de α, se obtiene un gráfico como el de la Figura 10.

Figura 10. *Puntos de inflexión identificados mediante los tres criterios utilizados en el conjunto de datos 3 (defunciones entre 2020 y 2023).*



Fuente. *Elaboración propia*.

El procesamiento del conjunto de datos permitió identificar 643 puntos de inflexión con el criterio 1 (azul), 691 puntos con el criterio 2 (rojo) y 47 con el criterio 3 (amarillo). Todos estos puntos calculados con un tamaño de máscara semanal (7) y un valor de 0.5 de α, los cuales se ubican en las fechas marcadas en el eje x. Sin embargo, este tipo de gráfica complica la identificación de los puntos de inflexión.

Para intentar describir la forma en que la propuesta de este trabajo detecta los puntos de inflexión, a continuación, se presenta una tabla con la información obtenida por la herramienta desarrollada, en la que se trabajan diferentes tamaños de máscara y de valores de α.

Tabla 2. *Puntos de inflexión identificados en el tercer conjunto de datos con diferentes tamaños de máscara y valores de α.*

Tamaño de la máscara	Criterio de ε	α = 0.1 (Estricto)	α = 0.5 (Mesurado)	α = 1 (Relajado)
7 (semanal)	1	218	643	843
	2	247	691	872
	3	2	47	98
31 (mensual)	1	281	601	825
	2	321	726	907
	3	5	25	83
365 (anual)	1	5	26	55
	2	7	36	74
	3	0	2	4

Fuente. *Elaboración propia*.

De la tabla anterior se puede deducir que el criterio 3 es el más exigente, pues, sin importar el tamaño de la máscara, ni del valor de α, el número de puntos de inflexión es menor que el de los otros dos criterios. Lo opuesto sucede con el criterio 2, que es el más permisivo,

al identificar un número mayor de puntos de inflexión, sin importar el tamaño de máscara, ni del valor de α.

Otra característica relevante de este procesamiento es que el tamaño de la máscara (el periodo analizado), influye directamente en el número de puntos de inflexión identificados. Para el caso de las defunciones diarias entre 2020 y 2023, entre más pequeña sea la máscara (semanal o quincenal), más puntos de inflexión se van a identificar, y si es mayor el tamaño de la máscara (semestral o anual), el número de puntos de inflexión identificados es menor.

A diferencia de los trabajos desarrollados en [10], [11], y [12], en las que se hace procesamiento sobre imágenes a nivel de pixeles, o bien entrenan modelos, para el reconocimiento de voz, en este trabajo la metodología se aplica directamente sobre el conjunto de datos, proporcionando un análisis más inmediato, sin requerir demasiados recursos computacionales.

CONCLUSIONES

La notación fue implementada a través de JavaScript en Google Apps Script satisfactoriamente, se desarrolló una interfaz HTML para introducir el tamaño de la máscara y el valor de α, se utilizó Google Sheets para cargar el conjunto de datos, guardar los resultados del procesamiento y ahí mismo se generan las gráficas correspondientes.

La pseudoderivada es una herramienta matemática útil para el reconocimiento de patrones. Tiene varias ventajas sobre otras herramientas de reconocimiento de patrones, incluyendo su simplicidad, robustez y eficiencia. Sin embargo, también tiene algunas desventajas, como su sensibilidad al ruido y la pérdida de información. El empleo de la pseudoderivada se comprobó que permite identificar un comportamiento en los datos como los valores máximos y mínimos, y en una segunda aplicación sobre los mismos, los puntos de inflexión y las alternancias entre las concavidades hacia abajo y hacia arriba.

La finalidad de identificar los puntos de inflexión radica en el hecho de reconocer si se trata de una concavidad hacia arriba o hacia abajo, pues de esta se puede comprobar que un fenómeno cambia su comportamiento al alza (concavidad hacia arriba), o bien, si su tendencia es a la baja (concavidad hacia abajo). La idea de automatizar este reconocimiento es evitar la inspección visual y que se identifiquen estos comportamientos.

El modelo permite obtener los puntos de inflexión de los conjuntos de datos de prueba, además permite manejar análisis en diferentes periodos (tamaño de máscara) y también considerando tres criterios para determinar la sensibilidad de identificación de los mismos (valor de α).

A diferencia de las aportaciones de [10], [11], y [12] en las que se hace procesamiento sobre imágenes y voz, lo que conlleva a un tratamiento computacionalmente más costoso, en cuanto a recursos y tiempo, la metodología propuesta en este trabajo permite manipular grandes cantidades de información de manera rápida y sin utilizar demasiados recursos, incluso se utilizó una herramienta basada en la nube (Google Sheets) para el procesamiento de los conjuntos de datos y generando los resultados mostrados.

A partir de este trabajo, se puede proponer como mejora, una estrategia para presentar los resultados obtenidos de manera intuitiva y fácil de interpretar, pues el modelo para trabajar con la pseudoderivada comprobó ser una buena alternativa para el reconocimiento de patrones, específicamente los puntos de inflexión. Además, de emplear otras herramientas para el tratamiento de datos como Python y sus librerías para mejorar la interpretación de los resultados, de una manera interactiva y clara.

Otra forma de mejorar la presentación de los resultados podría ser el empleo de los datos geográficos (en caso de tenerlos) mediante mapas de calor y paneles deslizantes interactivos (a manera de un dashboard básico), lo que puede representar una mejor experiencia del usuario.

AGRADECIMIENTOS

Los autores del presente trabajo agradecen a la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) y al Tecnológico Nacional de México (TecNM) por el apoyo brindado para la realización de esta investigación.

BIBLIOGRAFÍA

[1] Adrian, Cecilia, et al. “Theoretical Aspect In Formulating Assesment Model Of Big Data Analytics Environment.” Acta Mechanica Malaysia, vol. 1, no. 1, 2018, pp. 16–17., doi:10.26480/amm.01.2018.16.17.

[2] Zeng, Wenrong, et al. “Access Control for Big Data Using Data Content.” 2013 IEEE International Conference on Big Data, 2013, doi:10.1109/bigdata.2013.6691798.

[3] Fernández, Alicia, et al. “Pattern Recognition in Latin America in the ‘Big Data’ Era.” Pattern Recognition, vol. 48, no. 4, 2015, pp. 1185–1196., doi:10.1016/j.patcog.2014.04.012.

[4] Ianni, M., Masciari, E. & Sperlí, G. A survey of Big Data dimensions vs Social Networks analysis. J Intell Inf Syst 57, 73–100 (2021). https://doi.org/10.1007/s10844-020-00629-2

[5] Zheng, C., Zhang, Q., Long, G., Zhang, C., Young, S.D., Wang, W. (2020). Measuring time-sensitive and topic-specific influence in social networks with lstm and self-attention. IEEE Access, 8, 82481–82492.

[6] Tian, S., Mo, S., Wang, L., Peng, Z. (2020). Deep reinforcement learning-based approach to tackle topic-aware influence maximization. Data Science and Engineering, 1–11.

[7] Shao, H., Sun, D., Su, L., Wang, Z., Liu, D., Liu, S., Kaplan, L., Abdelzaher, T. (2020). Truth discovery with multi-modal data in social sensing. IEEE Transactions on Computers, 1–1.

[8] Franklinos, Lydia, et al. “Key Opportunities and Challenges for the Use of Big Data Inmigration Research and Policy.” 2020, doi:10.14324/111.444/000042.v1.

[9] Gaidarski, Ivan, and Pavlin Kutinchev. “Using Big Data for Data Leak Prevention.” 2019 Big Data, Knowledge and Control Systems Engineering

(BdKCSE), 2019, doi:10.1109/bdkcse48644.2019.9010596.
[10] Zhang, Y., Wang, J., & Liu, Y. (2023). Pseudoderivada-based feature extraction for image classification. Pattern Recognition, 132, 108925.
[11] Li, X., Zhang, H., & Sun, S. (2022). Speaker identification using pseudoderivatives. Speech Communication, 143, 109-118.
[12] Wang, Y., Zhang, X., & Li, X. (2021). Handwritten digit recognition using pseudoderivatives. International Journal of Pattern Recognition and Artificial Intelligence, 35(04), 2150017.
[13] Lohr, Sharon. "Big Data and Crime Statistics." Measuring Crime, 2019, pp. 125–136., doi:10.1201/9780429201189-10.
[14] Paz Grebe, M de la, Centeno, Angel M, Galazi, M Lucía, & Campos, M Soledad. (2021). Turning points: puntos de inflexión en la vida de los estudiantes de medicina. FEM: Revista de la Fundación Educación Médica, 24(6), 291-293. Epub 17 de enero de 2022.https://dx.doi.org/10.33588/fem.246.1157.
[15] Datos Abiertos Dirección General de Epidemiología, Influenza, COVID-19 y otros virus respiratorios [base de datos en línea]. Disponible en: https://www.gob.mx/salud/documentos/datos-abiertos-152127

Tabla de Roles de Contribución

Rol	Autor (es)
Conceptualización	Misael López Nava, Juan Reyes Reyes
Curación de datos	Misael López Nava
Metodología	Juan Reyes Reyes
Administración del proyecto	Misael López Nava, Juan Reyes Reyes
Supervisión	Juan Reyes Reyes, Carlos Manuel Astorga Zaragoza, Gloria Lilia Osorio Gordillo
Validación	Misael López Nava
Visualización	Luis José Muñiz Rascado
Redacción (borrador original)	Misael López Nava
Redacción (revisión y edición)	Juan Reyes Reyes