ISSN: 2594 - 2905

APPLICATION OF PSEUDODERIVATIVE FOR THE IDENTIFICATION OF INFLEXION POINTS IN DATASETS

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] National Technological Institute of Mexico/CENIDET,d21ce044@cenidet.tecnm.mx , https://orcid.org/0009-0005-0736-280X.
[2] National Technological Institute of Mexico/CENIDET,juan.rr@cenidet.tecnm.mx , https://orcid.org/0000-0001-7415-9987. ([3]) National Technological Institute of Mexico/CENIDET,gloria.og@cenidet.tecnm.mx , https://orcid.org/0000-0002-3445-8867. ([4]) National Technological Institute of Mexico/CENIDET,carlos.az@cenidet.tecnm.mx , https://orcid.org/0000-0002-4678-9006.
[5] National Technological Institute of Mexico/Zacatepec Technological Institute,luis.mr@zacatepec.tecnm.mx , https://orcid.org/0000-0003-2941-2482.

Abstract -- In the field of pattern recognition, there are techniques for identifying the desired characteristics of a data set, in other words, classifying information into predefined categories. The pseudoderivative is a tool that can be used in image recognition, speech recognition, handwriting recognition, etc. This paper 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 data to be processed at a specific moment based on the analysis of n immediately preceding and following data points, so that the number of elements for manipulation can be increased or decreased. The pseudo-derivative not only allows minimums and maximums to be identified in a first application, but also, when applied a second time to the same data set, allows the inflection points of the information on which it operates to be obtained.

Keywords: Big Data, Data Mining, Pattern Recognition.

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 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.

INTRODUCTION

The concept of Big Data refers to "massive" and often unstructured data, for which the processing capabilities of traditional data management tools are inadequate. Big Data can occupy terabytes and petabytes of storage space in various formats, including text, video, sound, images, and more. Although the term Big Data is relatively new, the trend of grouping and storing large volumes of information for future analysis is very old. [1] [2]

Big data is a set of data of great variety and formats, which accumulates in large volumes and at an ever-increasing speed. This is what is known as the 3 V's (dimensions) of Big Data. These dimensions are described below:

Volume. Organizations collect data from a wide variety of sources, including financial transactions, social media, sensors, and machine-to-machine communications. In the past, storage would have been a problem, but new technologies make the task easier.

Velocity. Data flows at an unprecedented rate and must therefore be managed in a timely manner. The increasing use of RFID (radio frequency identification) tags, sensors, and smart metering (meter reading systems) increases the need to manage data flows in real time or near real time.

Variety. Data comes in any format, from structured and numerical data in traditional databases to unstructured data, such as text documents, email, video, audio, quote data, and financial transactions. [3][4]

Received: June 12, 2025 | Accepted: September 13, 2025 | Published: October 8, 2025

text, email, video, audio, quote data, and financial transactions. [3][4]

However, with such rapid information generation, three other dimensions have been considered: variability, which refers to the exponential increase in the speed and variety of data combined with the fact that flows can be highly inconsistent and subject to periodic spikes; veracity, derived from the fact that it is extremely important to determine whether the data analyzed is relevant or not for the analysis of the information; and value, due to the ability of the data to generate useful and strategic information for organizations, enabling them to make better decisions, improve efficiency, and achieve greater business value. [5] [6]

Data mining studies methods and algorithms that enable the automatic extraction of synthesized information that allows the hidden relationships in large amounts of data to be characterized; it also aims to ensure that the information obtained has predictive capacity, facilitating efficient data analysis.[7] The term "data mining" encompasses various statistical and machine learning techniques focused primarily on the visualization, analysis, and modeling of information from massive databases. [8] [9]

A method for extracting image features based on pseudo-derivatives is proposed. Pseudo-derivatives are a generalization of traditional derivatives that can capture richer information about the local texture of an image. The proposed method uses pseudo-derivatives of different orders to calculate a set of features that are then used for image classification. The method is based on two main stages: 1) calculation of pseudoderivatives: Pseudoderivatives of different orders are calculated for each pixel in the image. 2) feature extraction: The pseudoderivatives are used to calculate a set of features that describe the local texture of the image. The proposed method was evaluated on a series of image datasets and compared with other feature extraction methods. The results showed that the proposed method achieves better classification accuracy in most cases. The feature extraction method based on pseudo-derivatives is a promising tool for image classification. [10]

The method is capable of capturing richer information about the local texture of an image than traditional derivatives, leading to better classification accuracy. The proposed method is computationally more expensive than other feature extraction methods. In addition, the performance of the method can be affected by noise in the images. The proposed method can be used in a variety of image classification applications, such as object recognition, scene classification, and defect detection.

Another method is proposed for speaker identification using pseudo-derivatives, which are features extracted from the speech signal that are robust to speech variations. The authors first extract pseudo-derivatives from the speech signal. They then use these pseudo-derivatives to train a speaker identification model. Finally, they evaluate the performance of the model on a speech database. The results show that the proposed method achieves speaker identification accuracy comparable to state-of-the-art methods. In addition, the proposed method is more robust to speech variations, such as noise and channel change. [11]

The proposed method is a promising alternative to traditional speaker identification methods. Its robustness to speech variations makes it suitable for a variety of applications, such as voice authentication and speech recognition. The study is based on a single speech database. Further research is needed to evaluate the performance of the proposed method on other speech databases. For this reason, this methodology can be used in a variety of applications, such as voice authentication, speech recognition, and speech synthesis.

Another method is proposed for handwritten digit recognition using pseudo-derivatives, which are local features that can be calculated quickly and efficiently and are robust to distortion and noise. The proposed method was evaluated on a database of handwritten digit images and compared with other existing methods. The results showed that the proposed method has comparable accuracy to existing methods, but with lower computational complexity. [12]

Furthermore, this methodology is a viable alternative for handwritten digit recognition. It is fast, efficient, and robust to distortion and noise, it is novel, and it has good potential for improving the accuracy of handwritten digit recognition. The experimental results are convincing and show that this method is comparable to existing methods. The article has some limitations, such as the lack of testing on other handwritten digit image databases. [13]

A turning point refers to two components: on the one hand, a change in direction (inflection) and a moment (point), i.e., they occur at a specific (chronological) time. [14]

They can be identified according to their positioning, i.e., downward concavity or upward concavity, which describes a behavior when a phenomenon from a specific moment tends to decrease (downward concavity), or when this phenomenon tends to increase (upward concavity). These behaviors are obtained from the analysis of data subsequent to and prior to a specific point in time. This interpretation highlights the importance of information analysis in this study, as it proposes its application to real data sets, in

Received: June 12, 2025 | Accepted: September 13, 2025 | Published: October 8, 2025

in this case, COVID-19 deaths from 2020 to 2023. [15]

DEVELOPMENT

Methodology.

The sub function is a function prior to the segmentation of an entire variable placed at a value k, and through this, information can be obtained from a k-th data point,
by averaging the previous and subsequent neighbors. This function is represented by equation 1.

$$sub_p(v,k) = [v(k - \frac{p-1}{2}), \ldots, v(k-1), v(k), v(k+1), \ldots, v(k + \frac{p-1}{2})] \quad (1)$$

Where:

v is each of the values to be used for processing.
k is the time instant of the data.

With this representation, an odd number of data can be analyzed, in proportion to the number of elements that make up the data set. In the case of a window of 5 data, it can be represented by equation 2.

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

In the segmentation stage, processing is performed to extract characteristics. For this work, the pseudo-derivative filter was used.
The pseudo-derivative filter can be represented by a vector such as that in equation 3, and when applied to a data set, it allows us to identify: when there is a zero crossing, a maximum or minimum is identified, and when it remains close to zero, it is constant.

$$z(k) = sub_5(b^I, k) \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \\ [1] \end{bmatrix}, k = 3, 4, \ldots, t_f - 2 \quad (3)$$

Where:

$z(k)$ is the first pseudo-derivative.
b_1^I is the data set to which the filter will be applied.
k is the time instant of the data.
tf-2 is the last value to be considered in the processing.

This function will allow maximum and minimum values to be identified when there is a zero crossing, as observed in the graph of the processing results.
In mathematics, an inflection point of a function is a point where the values of a continuous function in x change from one type of concavity to another. The curve crosses the tangent. The second derivative of the function f at the inflection point is zero, or does not exist. Based on this
Previously, the pseudo-derivative function can be applied a second time to the same data set for
obtain its inflection points. In this sense, the second operates on the vector resulting from the first application on the original data set.
Thus, the representation of the second pseudo-derivative would be as in equation 4.

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

Where:

$z'(k)$ is the second pseudo-derivative.
b_1^I is the data set to which the filter will be applied.
k is the time instant of the data.
tf-4 is the last value to be considered in the processing.

Once the feature vector with the
inflection points identified in the data set, a function is proposed that allows inflection points to be identified automatically. For this consideration, it is important to determine the appropriate size of the mask, as its size generates a more reliable approximation of the pattern (inflection point) in the analyzed data set.
A fundamental aspect for the formulation of an automated function that allows data processing without the need to obtain a visual (graphical) version and that determines a parameter to identify an inflection point, regardless of the number of samples in the data set. This value will be identified as ε, and will be obtained in order to discriminate the values obtained from the second pseudo-derivative, with the aim of providing an automation feature.
The function to identify that there is an inflection point at that point in the data set is shown in equation 5.

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

Where:

$g(k)$ is the function for identifying inflection points.
ε is the value used to discriminate between the values of $z'(k)$.

The aim of this function is to obtain a value of 1 to identify an inflection point and 0 when there is none.
Below are the functions proposed to discriminate the values, with different degrees of rigidity for detecting inflection points, using functions.
Criterion 1. One tenth of the difference between the maximum and minimum values. This is shown in equation 6.

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

This criterion identifies the inflection points without being so strict when the value of α is close to 1, that is, it tends to identify them without the analyzed values being so close to zero.

Received: June 12, 2025 | Accepted: September 13, 2025 | Published: October 8, 2025

Criterion 2. One tenth of the maximum difference between two contiguous values. It is shown in equation 7.

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

This criterion is stricter than the previous one.

Criterion 3. One hundredth of the maximum absolute magnitude of a value. It is shown in equation 8.

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

This criterion is the strictest for identifying inflection points.

Once the function g(k) is applied to each of the values obtained from the function z´(k), a vector of values 0 and 1 is generated, where the value 0 indicates that it is not an inflection point, while a 1 indicates that it is.

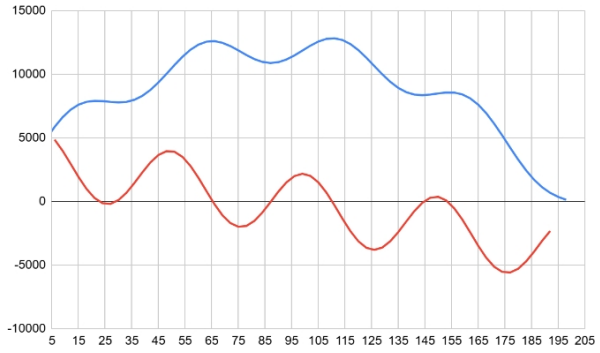
Application.

First test data set.

Considering a proposed data set consisting of 67 data points, taken in the same number of time units, except that the time units increase by three (0 to 198), the maximum and minimum values can be obtained by applying the pseudoderivative to the data set.

The following figure shows the graph with the blue line representing the values of the data set and the red line representing the behavior of the pseudo-derivative over the data set. It should be noted that the x-axis corresponds to the time units in which the samples were taken.

Figure 1. *Graph of the first pseudo-derivative*.



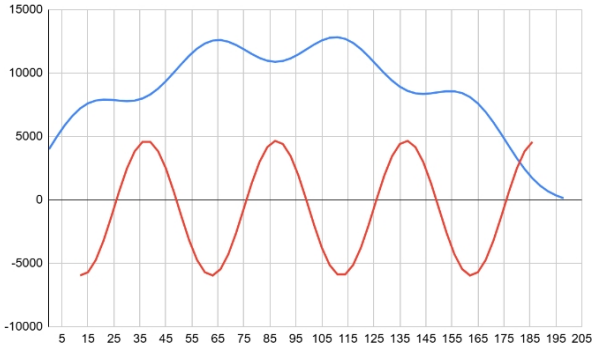
Source: *Own elaboration*.

Figure 1 shows the maximum and minimum values, which correspond to the zero crossings of the line (red) representing the pseudoderivative. The maximum values are identified at time values 21, 66, 111, and 153, while the minimum values are

located at times 30, 87, and 144.

When applying the second pseudo-derivative to the same data set, the behavior of the line representing it (red line) will now identify the inflection points at each zero crossing. This can be seen in Figure 2.

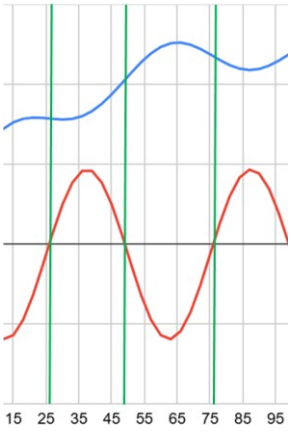
Figure 2. *Graph of the second pseudoderivative.*



Source: *Own elaboration*.

Figure 2 shows that the inflection points of the analyzed data are located at time values 27, 51, 78, 99, 126, 150, and 177. This can be verified in the fragment taken from Figure 2 and shown in Figure 3.

Figure 3. *Inflection point identified using the second pseudo-derivative.*



Source: *Own elaboration*.

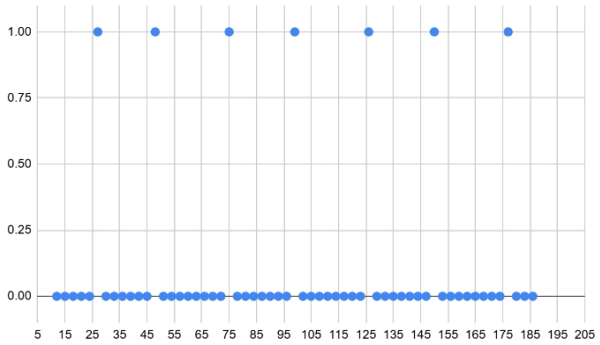
In this graph, the green lines confirm that, at each intersection of the pseudo-derivative (red line) with the x-axis, there is an inflection point in the actual data set (blue line), i.e., there is a change in the concavity of the line representing the behavior of the analyzed data. Figure 3 shows a change from negative to positive concavity at the first green line, and the opposite (a change from upward concavity to downward concavity), i.e., from positive to negative, at the second green line. A similar phenomenon to the first occurs at the third green line, where there is a change from negative to positive concavity.

This identification was made using functions generated by Google Sheets; however, it is important to highlight the importance of proposing a model for their automatic identification. For this purpose, the function identified as equation 5 will be used to discriminate the inflection points, and then the criteria identified as equations 6, 7, and 8 will be applied. This can be seen in Figure 4.

Received: June 12, 2025 | Accepted: September 13, 2025 | Published: October 8, 2025

ISSN: 2594 - 2905

Figure 4. *Identified inflection points.*



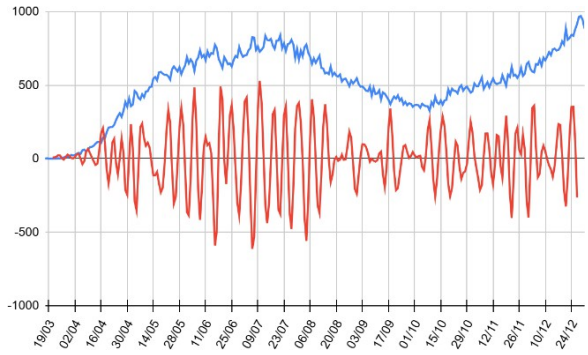
Source*: Own elaboration*.

Second set of test data.

Another analysis was performed on another dataset consisting of 290 samples sorted by the time they were obtained, corresponding to daily deaths from COVID-19 between March 17 and December 31, 2020.

The graph obtained to identify the behavior of the dataset and its second pseudo-derivative is shown in Figure 5.

Figure 5. *Inflection points of COVID-19 deaths during 2020.*



Source*: Own elaboration*.

The inflection points are identified at different dates in the data series and are visually identified at each zero crossing of the red line. The method of automatic identification, using equations 5 to 8, is shown below. The blue line represents daily deaths.

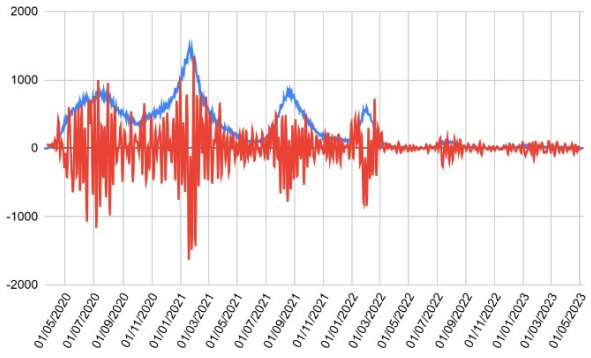
The disadvantage of this type of graphical representation of inflection points is that, as the amount of data increases, their identification becomes more difficult, accuracy is lost, and their interpretation becomes complicated.

Third test data set.

A third test set was used, corresponding to daily deaths from COVID-19 from March 17, 2020, to May 9, 2023, with a total of 1,149 data points, each of which is a daily sample.

When obtaining the graph of the second pseudo-derivative and comparing it with the behavior of the original set, it can be seen that, due to the amount of information being processed, the inflection points are not immediately apparent, as can be seen in Figure 6.

Figure 6. *Interface for identifying inflection points in COVID-19 deaths.*



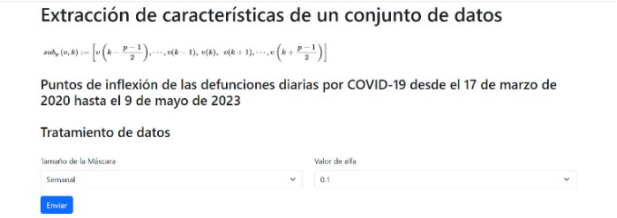
Source*: Own elaboration*.

In this regard, the visual identification of inflection points is very complicated, due to the behavior of the data being processed, which shows great variability in downward and upward alternations. Given the amount of data processed, the representation of the identified inflection points is inconvenient, highlighting that the proposed model does identify them.

For this case, the information was loaded into a Google Sheets spreadsheet, and the corresponding program was then developed in JavaScript, based on the algorithm designed to automate the functions described, using Google Apps Script, in which an HTML interface was also developed.

In this sense, this application works through a browser, in which the size of the mask is defined by the options: weekly, biweekly, monthly, quarterly, semi-annual, and annual; and the value of α (decimal values from 0 to 1, with intervals of 0.1). With this data, the values obtained from the pseudo-derivative and the second pseudo-derivative are generated, to subsequently generate the corresponding graph.

Figure 7. *Interface for identifying inflection points in COVID-19 deaths.*



Source*: Own elaboration*.

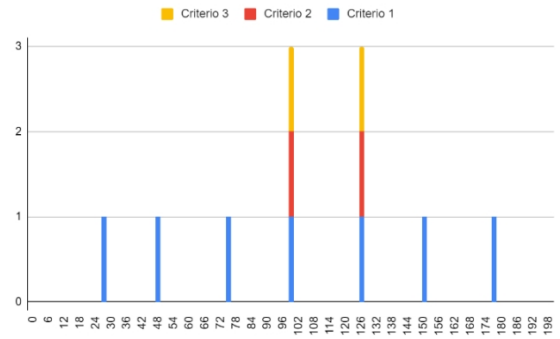
DISCUSSION AND ANALYSIS OF RESULTS

After applying equation 5 to identify the inflection points and the criteria defined as equations 6, 7, and 8, the graphs described below were obtained. From the first data set (67 values), a graph was obtained, which is shown in Figure 8.

Received: June 12, 2025 | Accepted: September 13, 2025 | Published: October 8, 2025

IPSUMTEC
ISSN: 2594 - 2905

Figure 8. *Inflection points identified using the three criteria used in data set 1.*

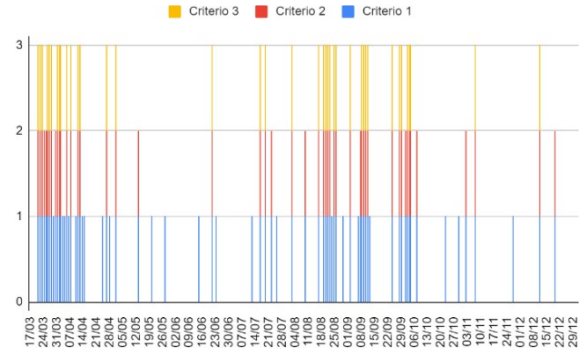


Source*: Own elaboration*.

It can be seen that criterion 1 (blue) identified seven inflection points, criterion 2 (red) identified only two, as did criterion 3 (yellow). In the graph, all these points are located at the times labeled on the x-axis.

From the second dataset (290 values), a graph was obtained, which is shown in Figure 9.

Figure 9. *Inflection points identified using the three criteria used in data set 2 (deaths during 2020).*



Source*: Own elaboration*.

Once each of the three criteria is applied to the values in the second data set, it can be seen that the number of inflection points increases compared to the first, which is directly proportional to the number of samples analyzed.

Sixty-nine inflection points were identified with criterion 1 (blue), 74 points with criterion 2 (red), and 2 with criterion 3 (yellow) in a weekly analysis, i.e., with a mask size (*sub* function*)* equal to 7, with a value of α=0.5 (neither strict nor relaxed). All these points are located on the dates marked on the x-axis. However, this type of graph complicates the identification of inflection points.

The following table shows the inflection points identified in the second dataset, with three different mask sizes 7, 31, and 93 (weekly, monthly, and quarterly, respectively), and with three different values of α 0.1, 0.5, and 1 (strict, moderate, and relaxed, respectively). This provides an overview of the behavior of the dataset.

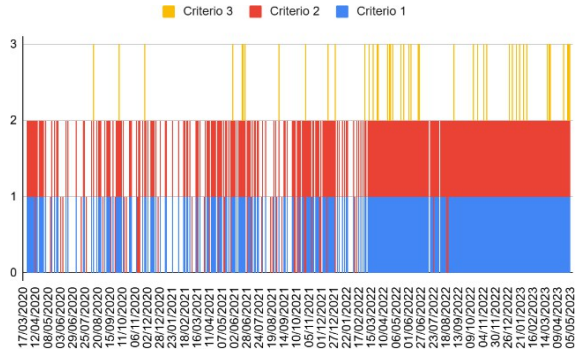
Table 1. *Inflection points identified in the second dataset with different mask sizes and α values.*

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

Source*: Own elaboration*.

With regard to the third dataset (1,149 values) used in the aforementioned proposal, depending on the mask size and the value of α, a graph such as the one in Figure 10 is obtained.

Figure 10. *Inflection points identified using the three criteria used in dataset 3 (deaths between 2020 and 2023).*



Source*: Own elaboration*.

Processing the dataset identified 643 inflection points with criterion 1 (blue), 691 points with criterion 2 (red), and 47 with criterion 3 (yellow). All these points were calculated with a weekly mask size (7) and an α value of 0.5, and are located on the dates marked on the x-axis. However, this type of graph complicates the identification of inflection points.

To try to describe how the proposal in this paper detects inflection points, below is a table with the information obtained by the tool developed, in which different mask sizes and α values are used.

Table 2. *Inflection points identified in the third data set with different mask sizes and α values.*

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

From the table above, it can be deduced that criterion 3 is the most demanding, since, regardless of the size of the mask or the value of α, the number of inflection points is lower than that of the other two criteria. The opposite is true for criterion 2, which is the most permissive,

Received: June 12, 2025 | Accepted: September 13, 2025 | Published: October 8, 2025

as it identifies a greater number of inflection points, regardless of the mask size or the value of α.

Another relevant feature of this processing is that

The size of the mask (the period analyzed) directly influences the number of inflection points identified. In the case of daily deaths between 2020 and 2023, the smaller the mask (weekly or biweekly), the more inflection points will be identified, and if the mask size is larger (semiannual or annual), the number of inflection points identified is lower. Unlike the work carried out in [10], [11], and [12], in which images are processed at the pixel level or models are trained for voice recognition, in this work the methodology is applied directly to the dataset, providing a more immediate analysis without requiring excessive computational resources.

CONCLUSIONS

The notation was successfully implemented using JavaScript in Google Apps Script. An HTML interface was developed to enter the mask size and the value of α. Google Sheets was used to load the dataset, save the processing results, and generate the corresponding graphs.

The pseudo-derivative is a useful mathematical tool for pattern recognition. It has several advantages over other pattern recognition tools, including its simplicity, robustness, and efficiency. However, it also has some disadvantages, such as its sensitivity to noise and loss of information. The use of the pseudoderivative was found to allow the identification of behavior in the data such as maximum and minimum values, and in a second application on the same data, inflection points and alternations between downward and upward concavities.

The purpose of identifying inflection points is to recognize whether it is an upward or downward concavity, as this can be used to verify whether a phenomenon is changing its behavior upward (upward concavity) or downward (downward concavity). The idea of automating this recognition is to avoid visual inspection and identify these behaviors.

The model allows the inflection points of the test data sets to be obtained. It also allows analyses to be performed over different periods (mask size) and considering three criteria to determine their identification sensitivity (α value). Unlike the contributions of [10], [11], and [12], which process images and voice, leading to more computationally expensive processing in terms of resources and time, the methodology proposed in this work allows large amounts of information to be manipulated quickly and without using too many resources. A cloud-based tool (Google Sheets) was even used to process the data sets and generating the results shown.

Based on this work, an improvement could be proposed: a strategy to present the results obtained in an intuitive and easy-to-interpret manner, as the model for working with the pseudo-derivative proved to be a good alternative for pattern recognition, specifically inflection points. In addition, other data processing tools such as Python and its libraries could be used to improve the interpretation of the results in an interactive and clear manner.

Another way to improve the presentation of the results could be to use geographic data (if available) through heat maps and interactive sliding panels (like a basic dashboard), which could provide a better user experience.

ACKNOWLEDGMENTS

The authors of this paper would like to thank the Secretariat of Science, Humanities, Technology, and Innovation (SECIHTI) and the National Technological Institute of Mexico (TecNM) for their support in carrying out this research.

BIBLIOGRAPHY

[1] Adrian, Cecilia, et al. "Theoretical Aspect In Formulating Assessment 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: inflection points in the lives of medical students. FEM: Journal of the Medical Education Foundation, 24(6), 291-293. Epub January 17, 2022.https://dx.doi.org/10.33588/fem.246.1157.
[15] Open Data General Directorate of Epidemiology, Influenza, COVID-19, and Other Respiratory Viruses [database online line]. Available at: https://www.gob.mx/salud/documentos/datos-abiertos-152127

Table of Contribution Roles

Role	Author(s)
Conceptualization	Misael López Nava, Juan Reyes Reyes
Data curation	Misael López Nava
Methodology	Juan Reyes Reyes
Administration of the project	Misael López Nava, Juan Reyes Reyes
Supervision	Juan Reyes Reyes, Carlos Manuel Astorga Zaragoza, Gloria Lilia Osorio Gordillo
Validation	Misael López Nava
Visualization	Luis José Muñiz Rascado
Drafting (original draft)	Misael López Nava
Writing (revision and editing)	Juan Reyes Reyes

Received: June 12, 2025 | Accepted: September 13, 2025 | Published: October 8, 2025