

# Aplicación de técnicas de reducción de dimensionalidad y visualización de datos en retroalimentaciones de cursos en línea

## Application of dimensionality reduction techniques and data visualization in online course feedback

Dorian Ruiz Alonso, Claudia Zepeda Cortés, Hilda Castillo Zacatelco, José Luis Carballido Carranza

Benemérita Universidad Autónoma de Puebla  
Facultad de Ciencias de la Computación  
[dorianr, czepedac, hildacz,]lcarballido7@gmail.com

Fecha de recepción: 14 de noviembre de 2023

Fecha de aceptación: 23 de abril de 2024

**Resumen.** El estudio versa sobre la integración de voluntades entre pequeños productores, Universidad de Colima, Cluster de las TI Colima A.C. y el INIFAP (Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias) mediante la intervención de un cultivo de plátano macho de 1.6 hectáreas sembradas en densidad media con aproximadamente 1800 plantas, en el municipio de Tecomán, Colima, México, mediante un procedimiento de análisis descriptivo, inductivo, experimental y longitudinal en el tiempo, se documentan tecnologías utilizadas en análisis de fertilidad del suelo y análisis satelital de cultivos. El estudio adquiere vital relevancia en el contexto de la seguridad alimentaria, el post covid y la presente guerra comercial de Rusia y Ucrania.

**Palabras clave:** Transformación digital, Clúster TI, Análisis satelital, Fertilidad de suelos, Seguridad alimentaria.

**Summary.** The study deals with the integration of wills among small producers, University of Colima, Cluster de las TI Colima A.C. and the INIFAP (National Institute of Forestry, Agriculture and Livestock Research) through the intervention of a plantain crop of 1.6 hectares planted in medium density with approximately 1800 plants, in the municipality of Tecomán, Colima, Mexico, through an analysis procedure descriptive, inductive, experimental and longitudinal in time, technologies used in soil fertility analysis and satellite analysis of crops are documented. The study acquires vital relevance in the context of food security, post-covid and the current trade war between Russia and Ukraine.

**Keywords:** Digital transformation, IT Cluster, Satellite analysis, Soil fertility, Food security.

**Resumen.** En este trabajo se explora la aplicabilidad de las técnicas de reducción de dimensionalidad para la visualización de retroalimentaciones generadas en cursos en línea. Se revisan las técnicas de reducción de dimensionalidad lineales y no lineales y sus posibles aplicaciones en el ámbito de los cursos en línea. La reducción de la dimensionalidad y visualización de las retroalimentaciones permite tener un mejor conocimiento de conjuntos de datos de gran dimensión para su posterior tratamiento en la creación de modelos agrupamiento.

**Palabras Clave:** Retroalimentación, reducción de dimensionalidad, e-learning.

**Summary.** This paper explores the applicability of dimensionality reduction techniques for the visualization of feedback generated in online courses. It is explored linear and nonlinear dimensionality reduction techniques and their applications in the field of online courses. Dimensionality reduction and visualization of feedbacks allow to have a better knowledge of high dimensional data sets for their subsequent treatment in the creation of clustering models.

**Keywords:** Feedback, dimensionality reduction, e-learning.

## 1 Introducción

En este trabajo se evalúan diversas técnicas de reducción de dimensionalidad y visualización con la finalidad de aplicarlos al análisis de retroalimentaciones generadas en cursos en línea.

La transformación de características es un método en la que se definen características nuevas como una representación de las que se encuentran en los conjuntos de datos originales. Entre estos métodos se encuentran la reducción de dimensionalidad, en la que los documentos son transformados en un nuevo espacio de características de menor dimensión, en donde las características son una combinación lineal de las características del conjunto de datos original (Aggarwal y Zahi, 2012).

El objetivo de la reducción de la dimensionalidad es encontrar representaciones de menor dimensión de datos que preservan las propiedades principales de un problema (Deng, et al, 2022). La reducción de la dimensionalidad permite el visualizar, comprimir y decrecer el tiempo de procesamiento de los datos. Tsai (2011), menciona que el proceso de reducir el número de dimensiones puede ser útil para distinguir las características o variables más importantes con el fin de proveer mejor conocimiento sobre la naturaleza de los datos, además de permitir detectar estructuras ocultas que permite el entendimiento y visualización de los datos.

Las técnicas de reducción de dimensionalidad pueden ser lineales como el análisis de componentes principales (PCA) y Escalamiento multidimensional (MDS) o no lineales como el embebido lineal local (LLE) y el mapeo de características isométricas (ISOMAP).

Este trabajo se aplican técnicas de reducción de dimensionalidad lineal y no lineal para visualizar retroalimentaciones generadas en cursos en línea con el fin de tener un mejor panorama del conjunto de datos para su posterior tratamiento en la creación de modelos de agrupación.

Aunque existen trabajos que aplican técnicas de reducción de dimensionalidad en casos como el análisis de blogs (Tsai, 2011), diagnóstico cáncer (Galarza, 2017), imágenes satelitales (Chacón, 2021), de lo revisado en la literatura las técnicas no se han aplicado para visualizar retroalimentaciones de cursos en línea.

Las retroalimentaciones se han identificado ser un factor importante en el aprendizaje de los estudiantes (Rivera, 2021), por lo que encontrar formas que permitan analizarlas de manera automática es relevante.

En los apartados siguientes se presentan una descripción de las técnicas de reducción de dimensionalidad lineales y no lineales, metodología, experimentos, resultados y conclusiones.

## **2 Técnicas de reducción de dimensionalidad lineal y no lineal**

El propósito de las técnicas de reducción de dimensionalidad es reemplazar las variables originales por un conjunto menor de variables subyacentes (Tsai, 2011). Las técnicas pueden clasificarse en lineales y no lineales.

### **2.1 Análisis de componentes principales (PCA)**

El análisis de componentes principales es una técnica de reducción de dimensionalidad lineal que busca conseguir un nuevo conjunto de variables o componentes principales que están interrelacionadas entre sí a través de transformaciones ortogonales realizadas al conjunto de variables originales (Valencia, 2019).

PCA proyecta datos de  $n$ -dimensiones en un subespacio de  $d$ -dimensiones de menor tamaño en una forma que minimiza el error de suma cuadrática, o maximiza la varianza, o da distribuciones proyectadas no correlacionadas.

### **2.2 Escalado multidimensional (MDS)**

La técnica lineal escalado multidimensional (MDS) es un enfoque que busca realizar una representación de datos de menor dimensionalidad, mientras trata de preservar las distancias entre los puntos. Es un método para representar una matriz de disimilaridad en un cierto número de dimensiones. El método trata de preservar la estructura de los datos originales en el espacio de destino, de manera que las relaciones entre los puntos se mantengan lo más cercanas posible a las relaciones originales (Bravo, 2020).

### **2.3 Análisis semántico latente (LSA)**

El análisis semántico latente (LSA) es una técnica lineal de recuperación de información y clasificación, utiliza una matriz término-documento que describe patrones de distribución de términos (palabras) a través de un conjunto de documentos, después encuentra una aproximación de bajo orden y menor ruido que la matriz original.

### **2.4 Mapeo de características isométricas (ISOMAP)**

La técnica de reducción de dimensionalidad no lineal ISOMAP es un mapeo que preserva una medida de distancia definida sobre una variedad de dimensión inferior que puede ajustarse a los datos razonablemente bien.

Asume que aprender esta variedad es la clave para tener éxito en el análisis de datos, y las distancias entre los puntos son intrínsecas o distancias geodésicas entre los puntos de la variedad (Deng, 2022).

Puede considerarse como una extensión de MDS o PCA. ISOMAP (Tanenbaum, 2000).

## **2.5 Embebido lineal local (LLE)**

La técnica embebido lineal local, busca hallar la variedad de baja dimensión dentro del conjunto de datos a trabajar los cuales son de alta dimensión y conservar la propiedad de que dos puntos que estaban próximos en el espacio origen lo seguirán estando en el nuevo espacio de dimensión realizando una búsqueda de vecinos más cercanos para construir la matriz de peso y descomponer parcialmente los valores propios (Quansheng et al, 2008).

## **2.6 T-SNE**

La técnica no lineal T-SNE, Rengifo y Gallego (2016), mencionan que es una distribución estocástica de puntos más cercanos, este método hace uso de una distribución que mide la similitud entre pares de objetos de entradas y una medida de distribución entre parejas similares de los correspondientes puntos de análisis, representando así un conjunto con una menor dimensión. Este método comienza por encontrar patrones en los datos mediante la identificación de grupos que han sido observados según la similitud de los puntos de datos con varias características.

# **3 Metodología**

## **3.1 Datos**

Se utilizó del conjunto de datos multi-etiqueta utilizado en Ruiz (et al 2011) que consiste en 5007 retroalimentaciones generadas por docentes en cursos en línea de la facultad de derecho de una universidad pública de México. El conjunto de datos multi-etiqueta cuenta con retroalimentaciones que fueron etiquetadas manualmente por expertos quienes utilizaron el modelo de Hattie y Timperley (2007) el cual distingue los niveles tarea (FT), proceso (FP), regulación (FR), y elogios (FS) para realizar el análisis de efectividad de los comentarios.

## **3.2 Extracción de características**

El conjunto de datos se preprocesó a través de paso de retroalimentaciones a minúsculas, lematización, corrección ortográfica, stop words, signos diacríticos y truncamiento. Después se extrajeron las características utilizando un-gramas como unidad de separación de textos y se transformaron las retroalimentaciones en una matriz término-documento calculando el valor TF-IDF.

## **3.3 Aplicación de técnicas de reducción de dimensionalidad**

La matriz TF-IDF fue la entrada de cada una de las técnicas de reducción de dimensionalidad lineal y no lineal PCA, MDS, LSA, ISOMAP, LLE y T-SNE para generar una nueva matriz de 2 características.

## **3.4 Visualización**

Las matrices de dos dimensiones se utilizaron para la visualización de las retroalimentaciones mediante el uso de gráficas de dispersión que muestran cada una en un espacio de dos dimensiones. Por cada método de reducción de dimensionalidad, se obtuvieron 6 gráficos, 5 corresponden a la separación de las etiquetas utilizando relevancia binaria y el último a la generación del conjunto potencia mostrando solo aquellas retroalimentaciones que pertenecen solamente a uno de los niveles.

## 4 Experimentos y Resultados

Las gráficas de la Figura 1 visualizan las retroalimentaciones después de la aplicación del método de reducción de dimensionalidad lineal PCA. Se muestra que el método permite discriminar las retroalimentaciones que están en el nivel proceso (FP) (en color azul) de las que no lo están (color rojo) al mostrarlas agrupadas en lados opuestos, aunque solapamiento entre los grupos. Para los niveles FS, FR y FO se muestra alto grado de solapamiento.

Al mostrar las retroalimentaciones que pertenecen a un solo nivel se logran notar una discriminación de los comentarios que tienen las etiquetas únicas FT (moradas), FP (azul marino) y FS (azul claro).

La técnica de reducción de dimensionalidad lineal MDS Figura 2, realiza una mejor discriminación de los comentarios en el nivel FP, aunque hay solapamiento al mostrar dos grupos separadas. Al discriminar comentarios utilizando la transformación LP, muestra una alta discriminación entre FT y FP de igual forma una buena discriminación entre FP y FS. Lo que podría dar indicios de que las retroalimentaciones enfocadas al proceso no están combinadas con elogios a estudiantes.

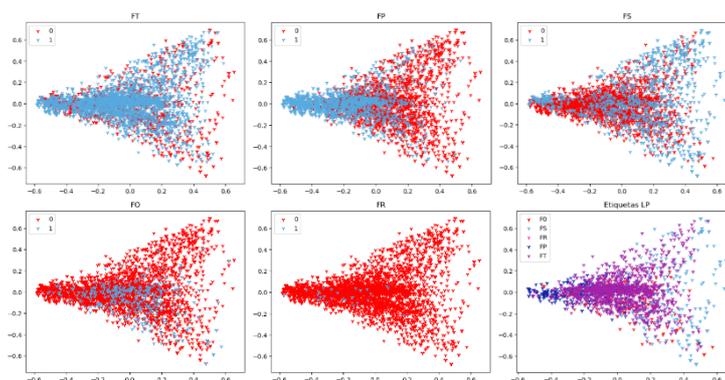


Figura 1. Visualización de retroalimentaciones con características PCA

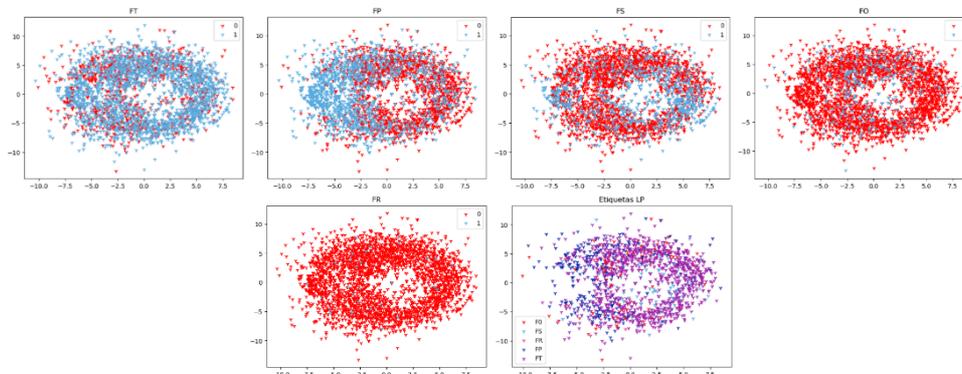
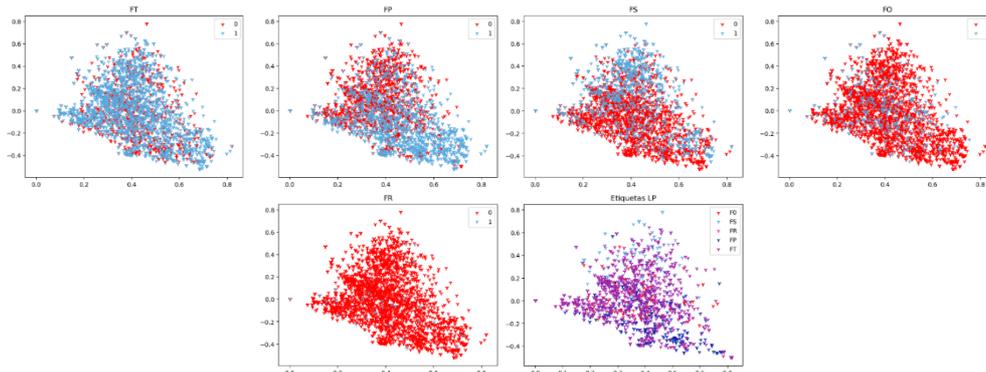


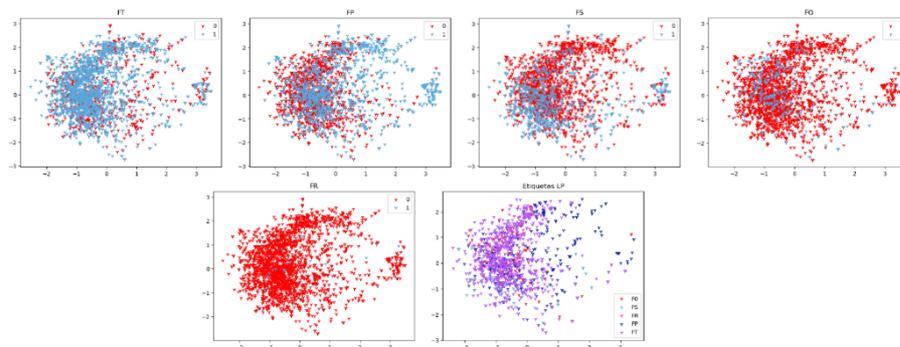
Figura 2. Visualización de retroalimentaciones con características MDS

La técnica de reducción de dimensionalidad lineal LSA Figura 3, discrimina de mejor manera comentarios de categoría FP y FT. En la discriminación de comentarios con etiqueta única lo hace de mejor manera para FP de los FS. De igual forma los FS y de los FO.



**Figura 3.** Visualización de retroalimentaciones con características LSA

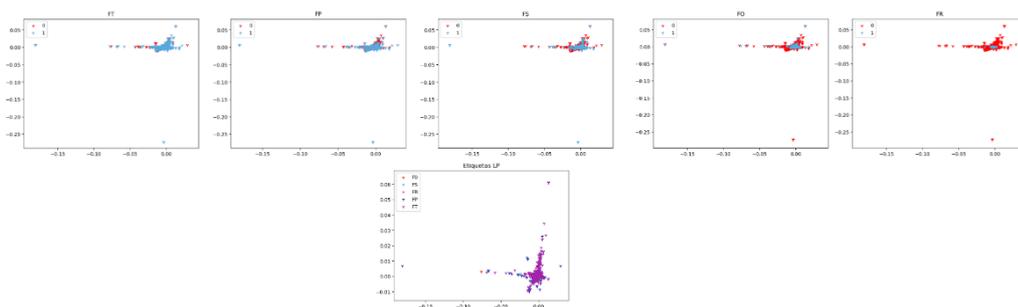
La técnica de reducción de dimensionalidad no lineal ISOMAP Figura 4 muestra un bajo desempeño en discriminar los comentarios al estudiante para identificar si encuentra en un nivel específico o si se encuentra o en alguna de las clases. De igual forma solapa los comentarios etiquetados con una sola etiqueta.



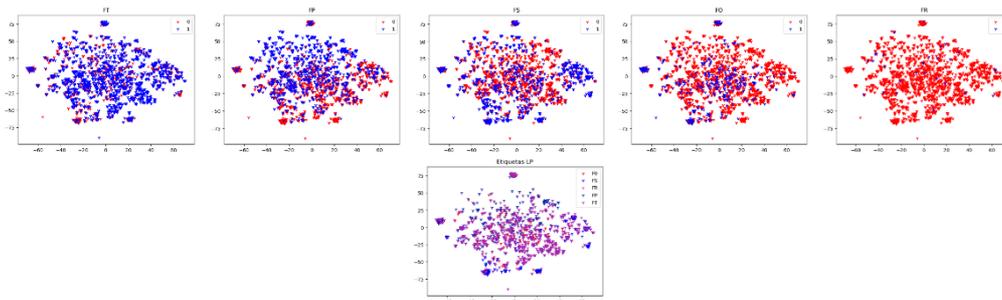
**Figura 4.** Visualización de retroalimentaciones con características ISOMAP

La técnica de reducción de dimensionalidad no lineal LLE Figura 5, no muestra un buen desempeño en la discriminación de los comentarios para ningún nivel. Cabe destacar que localiza comentarios que se pueden considerar ruido por la lejanía en la que se ubican entre las demás muestras.

La técnica de reducción de dimensionalidad no lineal T-SNE Figura 6, muestra un alto grado de solapamiento para todas las categorías.



**Figura 5.** Visualización de retroalimentaciones con características LLE



**Figura 6.** Gráficas de dispersión con características T-SNE.

## 5 Conclusiones y Trabajo Futuro

La aplicación de técnicas de reducción de dimensionalidad de retroalimentaciones en cursos en línea y visualización se ha evaluado en este trabajo. Se utilizaron técnicas de reducción de dimensionalidad lineal y no lineal. Se muestra que las técnicas de reducción de dimensionalidad lineales tienen un mejor desempeño que las no lineales. Aunque ningún método logra un buen desempeño para discriminar todas las categorías si lo realizan para separar algunas lo que puede resultar de utilidad para entender la forma en que se retroalimenta, por ejemplo, se puede destacar que cuando se enfocan a la tarea no hay elogios.

Aplicar las técnicas de reducción de dimensionalidad y visualización permite entender mejor el problema y descubrir relaciones entre las variables. Como trabajo futuro se analizará como responden la reducción de dimensionalidad en el entrenamiento de modelos de aprendizaje de máquina no supervisados.

## Referencias

- Aggarwal, C. C., & Zhai, C. (2012). "A survey of text clustering algorithms." *Mining text data*, 77-128.
- Bravo, A. (2020). "Reducción de dimensiones: revisión y aplicaciones en clasificación automática".
- Chacón, J. M. (2021). "Reducción de dimensionalidad aplicada a imágenes de satélite", (Bachelor's thesis).
- Cunningham, J. P., & Ghahramani, Z. (2015). "Linear dimensionality reduction: Survey, insights, and generalizations." *The Journal of Machine Learning Research*, 16(1), 2859-2900.
- Deng, L. Y., Garzon, M., & Kumar, N. (2022). "What Is Dimensionality Reduction (DR)?." In *Dimensionality Reduction in Data Science* (pp. 67-77). Cham: Springer International Publishing.
- Galarza Hernández, J. (2017). "Reducción de dimensionalidad en Machine Learning. Diagnóstico de cáncer de mama bsado en datos genómicos y de imagen", (Doctoral dissertation, Universitat Politècnica de València).
- Hattie, J., & Timperley, H. (2007). "The Power of Feedback." *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Rengifo Almanza, D. F., & Gallego Rincón, J. F. (2016). "Comparación de técnicas de reducción de dimensionalidad para la clasificación de actividades físicas humanas utilizando métodos estadísticos".
- Rivero, V. M. H., Bonilla, P. J. S., & Alonso, J. J. S. (2021). "Feedback y autorregulación del aprendizaje en educación superior". *Revista de Investigación Educativa*, 39(1), 227-248.
- Ruiz Alonso, D., Zepeda Cortés, C., Castillo Zacatelco, H., Carballido Carranza, J. L., & García Cué, J. L. (2022). "Multi-label classification of feedbacks." *Journal of Intelligent & Fuzzy Systems*, 42(5), 4337-4343.
- Tenenbaum, J. B., Silva, V. D., & Langford, J. C. (2000). "A global geometric framework for nonlinear dimensionality reduction." *science*, 290(5500), 2319-2323.
- Tsai, F. S. (2011). "Dimensionality reduction techniques for blog visualization." *Expert Systems with Applications*, 38(3), 2766–2773.
- Valencia Guerrón, M. A. (2019). "Comparación de métodos de reducción de dimensionalidad enfocados a algoritmos de clasificación supervisados aplicado a datos de redes de sensores" (Bachelor's thesis).