

Evaluación Sísmica del Tren Interurbano México-Toluca con simulaciones de carga viva y Ciencia de Datos

Seismic Evaluation of the Mexico-Toluca Interurban Train with Live Load Simulations and Data Science

Valdovinos-Rosas, R.M.^{1*} , Arellano-Magdaleno, O. D.¹ , López-Díaz, J. L.¹ , Delgado-Hernández, D. J.¹ 
(*rvaldovinosr@uaemex.mx)

Recibido: 6 de junio de 2024 Aceptado: 19 de junio de 2024

RESUMEN

El Gobierno Federal de México está construyendo el Tren Interurbano México-Toluca para abordar el desafío de transportar a más de 230 mil personas diariamente entre las ciudades de México y Toluca. Para evaluar la construcción del tren se han propuesto estudios, de los que la mayoría se realizan a base de simulaciones por la poca disposición de datos reales. En este artículo se realiza el análisis del modelo estructural del tren interurbano México-Toluca al ser sometido a un terremoto de magnitud similar al del 22 de mayo de 1960 en Chile y del 19 de septiembre de 1985 en México considerando simulaciones ejecutadas

con el programa SAP 2000, mientras el tren se encuentra en movimiento a una velocidad determinada y con carga máxima de personas. Para la investigación se utiliza el lenguaje de programación Python y el software libre WEKA siguiendo el proceso general de ciencia y analítica de datos. El análisis de clúster mostró un severo traslape de clases en los datos del sismo en México. La clasificación mejoró significativamente al usar conjuntos derivados del agrupamiento con el algoritmo EM.

Palabras clave: *Tren interurbano, Sismo, Ciencia de datos, Agrupamiento, Machine Learning.*

¹Facultad de Ingeniería de la Universidad Autónoma del Estado de México. Paseo Universidad S/N. Toluca, Estado de México, México. CP. 50130.



ABSTRACT

The Federal Government of Mexico is constructing the Mexico-Toluca Interurban Train to address the challenge of transporting over 230,000 people daily between the cities of Mexico and Toluca. To evaluate the train's construction, various studies have been proposed, most of which rely on simulations due to the limited availability of real data. This article analyzes the structural model of the Mexico-Toluca interurban train when subjected to an earthquake of similar magnitude to the ones on May 22, 1960, in Chile and September 19, 1985, in Mexico. The simulations are executed using the SAP 2000 program while the train is in motion at a specific speed and with maximum passenger load. The research utilizes Python programming language and the open-source software WEKA, following the general data science and analytics process. The cluster analysis revealed a severe overlap of classes in the Mexico earthquake data. Classification significantly improved when using sets derived from the clustering with the EM algorithm.

Keywords: *Interurban train, Earthquake, Data science, Clustering, Machine learning.*

INTRODUCCIÓN

El Tren Interurbano México-Toluca es un proyecto del Gobierno Federal que busca conectar la capital del país con la ciudad de Toluca en un tiempo de viaje de 40 minutos. Este medio de transporte tendrá una longitud total de 57.7 km y contará con 7 estaciones (2 terminales y 5 intermedias) [1]. El análisis de viaductos elevados y puentes es importante, ya que pueden afectar estructuras como terremotos, sobrecargas, lluvia en exceso, falta de mantenimiento, plasticidad, fatiga, fractura, etc. Estas situaciones pueden provocar consecuencias negativas en las estructuras y desencadenar eventos no deseados, por ello, para evaluar su seguridad se analiza la estructura de la obra.

México, está ubicado en la placa tectónica norteamericana y limitando al sur y oeste con las placas de Cocos, Rivera y del Pacífico, se encuentra en el “Cinturón de Fuego”, una zona sísmicamente activa a nivel mundial. Esta situación motiva el estudio presentado en este artículo, evaluando la respuesta del tren Interurbano ante un terremoto mientras está en movimiento y transportando el máximo número de pasajeros (711 personas sentadas y de pie). Se estudiará una posible condición crítica de carga antes de la operación real para garantizar un comportamiento seguro tras la inauguración de la obra.

En específico se analizan dos terremotos: el del 22 de mayo de 1960 en Chile (9.5 Mw) correspondiente al más fuerte sismo registrado en el planeta, y el del 19 de septiembre de 1985 en México (8.1 Mw), significativo por la cercanía del epicentro con la construcción y su impacto en el país. Para este fin, se hace uso de técnicas de Ciencia de datos, como los algoritmos de agrupamiento y los de aprendizaje automático.

ESTADO DEL ARTE

Thai [2] realiza una revisión exhaustiva de las aplicaciones del Machine Learning en la ingeniería estructural, en el monitoreo de la salud estructural y en el diseño de estructuras de concreto. Además, ese autor presenta hallazgos, desafíos, recomendaciones y proporciona bases de datos y códigos de varios algoritmos como referencia.

Mangalathu S. y Jeon [3] utilizaron métodos de Machine Learning para predecir el modo de falla en columnas de puentes de concreto armado circular, con 311 muestras experimentales y tres tipos de fallas consideradas. Los métodos evaluados incluyeron análisis discriminante cuadrático, regla del vecino más cercano, árboles de decisión, bosques aleatorios, Naive Bayes y redes neuronales artificiales. Los resultados mostraron que las redes neuronales artificiales lograron una precisión del 91 % en la detección de fallas, superando a los otros métodos.

Nguyen et al. [4] presentaron un método para detectar daños en puentes usando cambios en la curvatura modal y Convolutional Neural Networks (CNN). El Gapped Smoothing Method (GSM) identifica la ubicación de los daños sin datos de referencia de estructuras sanas. Utilizando el puente Bonghi como ejemplo, se entrenó una CNN con imágenes del índice de daños del GSM para clasificar la ubicación del daño en una viga numérica. Los resultados muestran que la combinación de GSM y CNN puede detectar y localizar daños en puentes.

Mahmoudi et al. [5] realizan la detección de daños en edificios de muros de corte de concreto durante movimientos sísmicos. Utilizaron el algoritmo *K*-Nearest Neighbor ajustado con optimización bayesiana como modelo de predicción. Evaluaron el marco en tres edificios de referencia con datos generalizados, mostrando su eficacia en la identificación precisa de daños.

En el estudio de Abdelijaber et. al. [6] localizan daños estructurales en tiempo real utilizando redes neuronales convolucionales unidimensionales. Este enfoque destaca por su capacidad para extraer automáticamente características óptimas de las señales de aceleración y su eficiencia computacional como se demostró en experimentos realizados en un simulador de graderío.

En el estudio de Flah et al. [7] los algoritmos de Aprendizaje Automático (ML) en el Monitoreo de la Salud Estructural (SHM) son cada vez más relevantes por su capacidad para detectar daños en estructuras civiles. Con el crecimiento del Internet de las Cosas y Big data, estas aplicaciones se expandirán rápidamente. Esta revisión clasifica los algoritmos de ML en SHM basado en vibraciones e imágenes, analiza su eficacia y ofrece recomendaciones, destacando las brechas de conocimiento y las necesidades futuras de investigación.

Sarothi et al. [8] exploraron los modos de falla en conexiones atornilladas de doble corte en acero

estructural usando varios algoritmos en una base de datos de 455 resultados experimentales. Algoritmos como Random Forest, CatBoost, XGBoost y Gradient Boosting demostraron una precisión del 90-92 % en la clasificación de modos de falla, destacando la importancia de la relación entre la distancia al borde y el diámetro del perno en el rendimiento de estas conexiones.

Abdellatif et al. [9] identificaron grietas en imágenes de ingeniería civil mediante el algoritmo CrackIT, que ofrece enfoques basados en bloques y en píxeles. Proponen una nueva metodología que combina ambos para una detección precisa y tolerante al ruido. Evaluaron este enfoque en tres conjuntos de datos de imágenes de grietas, comparándolo con CrackIT y tres métodos de aprendizaje profundo. Su propuesta superó a los métodos existentes, reduciendo errores de discretización y mejorando la localización de grietas.

En línea con las investigaciones descritas, se realizó un estudio de dos fases para analizar la respuesta del tren interurbano México-Toluca frente a terremotos de gran magnitud y cargas excesivas, con el objetivo de evaluar su confiabilidad y seguridad para los futuros usuarios.

MÉTODOS DE APRENDIZAJE AUTOMÁTICO

Tal como se mencionó anteriormente, el estudio presentado en este artículo es en dos sentidos:

1. Descriptivo: Se analiza la estructura de los datos sintéticos previamente clasificados aplicando algoritmos de agrupamiento, para conocer la calidad de la distribución de datos: K-medias y EM
2. Predictivo: Con el agrupamiento obtenido se entrenaron los modelos de ML: Perceptrón Multicapa, Support Vector Machine y Árbol de Decisión para analizar su predicción.

Algoritmo K-medias

K-medias es un algoritmo de agrupamiento para lo que se requiere la especificación de una semilla k que determina la cantidad de grupos a obtener. El proceso general es el siguiente [10]:

1. Seleccionar los k centros iniciales.
2. Asignar las instancias (x) del conjunto de datos a su centroide más cercano (c) utilizando una métrica de similitud, que comúnmente es la distancia euclídea, Ec. (1).

$$d_{x,c} = \sqrt{\sum (x - c)^2} \quad (1)$$

3. Recalcular los centros considerando las instancias asignadas a ellos, Ec. (2).

$$\overline{c_j} = \frac{1}{c_j} \sum_{i=1}^{c_j} x_{ij} \quad (2)$$

4. Repetir los pasos 2 y 3 hasta que los centroides convergen.

Algoritmo EM

Expectation-Maximitation (EM) es un algoritmo no supervisado que identifica las distribuciones gaussianas de los datos mediante la generación de grupos considerando los siguientes pasos [11] :

1. Inicialización aleatoria del número de grupos esperados, así como los parámetros de las distribuciones gaussianas, considerando el rango de valores posibles de los datos.
2. Expectation. Realizar el cálculo de las distancias de los datos con respecto a las medias de las Gaussianas para determinar la distribución de probabilidad de los datos.
3. Maximitation. La distribución de probabilidad obtenida en la etapa anterior, son modificados mediante el cálculo de las nuevas medias y matrices de covarianza.
4. Repetir los pasos 1 y 2 hasta que no haya cambios en la asignación de datos a los grupos.

Perceptrón Multicapa

El Perceptrón multicapa (PM) es un modelo de clasificación supervisada popular debido a su robustez. Su topología consiste en tres tipos de capas: de entrada, oculta y de salida (ver Figura 1).

El algoritmo más utilizado para entrenar el PM es el Backpropagation, que incluye dos etapas principales: propagación hacia adelante y propagación hacia atrás de la señal [12]. El proceso general es el siguiente:

1. Inicialización de los pesos de las conexiones y los parámetros libres de la red (la razón de aprendizaje, factor momento, mínimo error esperado, entre otros).
2. Propagación hacia adelante: Cada instancia del conjunto de datos se presenta a la red y se propaga la señal hasta la capa de salida, donde se determina el error respecto a la salida esperada.

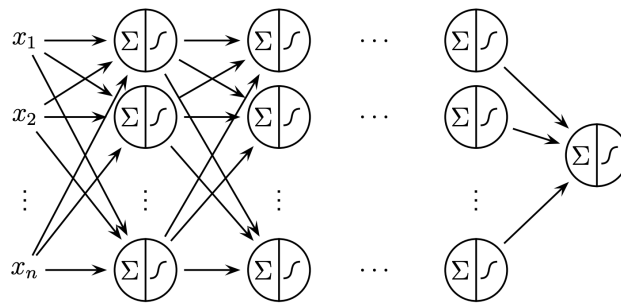


FIGURA 1. Topología típica de un PM. $[x_1 \dots x_n]$ representan el vector de entrada, Σ Representa la regla de propagación. f Denota la función de activación.

3. Propagación hacia atrás: La señal se propaga de la capa de salida y hasta la capa de entrada realizando el ajuste de los pesos considerando el gradiente descendente para minimizar el error cuadrático medio observado con cada instancia.
4. Los pasos 2 y 3 son repetidos hasta la convergencia de la red a un mínimo error obtenido.

Support Vector Machine

El objetivo del *Support Vector Machine* (SVM) es identificar el hiperplano que separa de forma óptima las clases de un problema, que tenga en consideración maximizar la distancia existente entre éste y las instancias más cercanas a él (ver Figura 2).

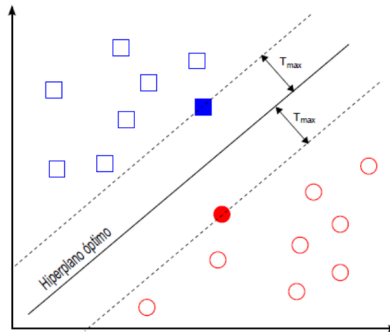


FIGURA 2. Hiperplano óptimo en un problema de dos clases.

El proceso que sigue el SVM consta de los siguientes pasos [13]:

1. Realizar el mapeo de los datos. En caso de ser necesario, transformar los datos de entrada a un espacio de dimensión alta (truco de kernel).
2. Identificar el hiperplano que mejor separa las clases en el problema.

3. Encontrar los vectores de soporte, que corresponden a dos hiperplanos paralelos al hiperplano que separa las clases, maximizando la distancia a la instancia más cercana a él.

Árbol de Decisión

El árbol de decisión es un clasificador supervisado que realiza particiones repetidas en el conjunto de datos para agrupar elementos de una sola clase. Su construcción sigue cuatro pasos.

1. Para cada atributo calcular la entropía y la ganancia de información normalizada.
2. Tomar como padre al atributo que tenga la ganancia más alta y crear un nodo de decisión que lo divida considerando los valores que el atributo adquiere.
3. Repetir en las sablistas obtenidas por división y agregar estos nodos como hijos.
4. Finalizar cuando se llegue a las hojas.

Se utilizó el lenguaje de programación Python con ayuda de bibliotecas fundamentales para el análisis de datos, la manipulación y la visualización, como lo son pandas, numpy y matplotlib.

METODOLOGÍA

Para dar solución al problema planteado se siguió proceso presentado en la Figura 3.

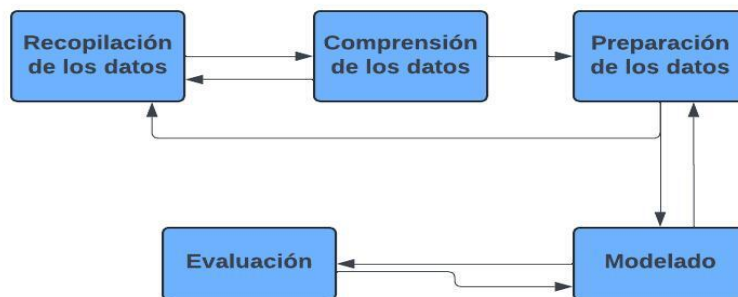


FIGURA 3. Estrategia de solución utilizada.

Compresión de los datos

Utilizando como valores de referencia los relacionados a la estructura física del tren interurbano México-Toluca [14], en un estudio previo [15] se diseñó un modelo estructural utilizando el programa Structural Analysis Program (SAP 2000). A partir de este modelo se generaron distintas simulaciones considerando

los sismos de Chile y de México respectivamente con el fin de observar la variabilidad y el comportamiento del conjunto de variables. Esto dio paso a la construcción de una Red Bayesiana con 14 nodos, que relaciona a las variables de sismo y carga viva máxima, con los elementos mecánicos de una pila de soporte y sus desplazamientos lineales y angulares. Después se ajustó la red bayesiana propuesta inicialmente modificando la variable relacionada con los sismos para considerar movimientos telúricos más intensos, para someter la estructura a un mayor esfuerzo y determinar la probabilidad de falla.

Para ambos sismos, se consideraron las características bajo el criterio de las variables del sismo, la carga viva máxima y los elementos mecánicos de una pila de apoyo que es donde normalmente ocurren las fallas de un viaducto. En el estudio, las 7 características del conjunto de datos son: factor de amplificación del sismo (GF), peso total de los pasajeros más peso de los ejes del tren (Wt), máximo esfuerzo por carga axial (P), máximo esfuerzo cortante en dirección X y Y (V_x , V_y), máximo momento de flexión respecto a X y Y (M_{x1} , M_{y1}), máximo momento de torsión (T), desplazamientos lineales en U_x , U_y , U_z , y giros con respecto a los tres ejes R_x , R_y , R_z (Figura 4).

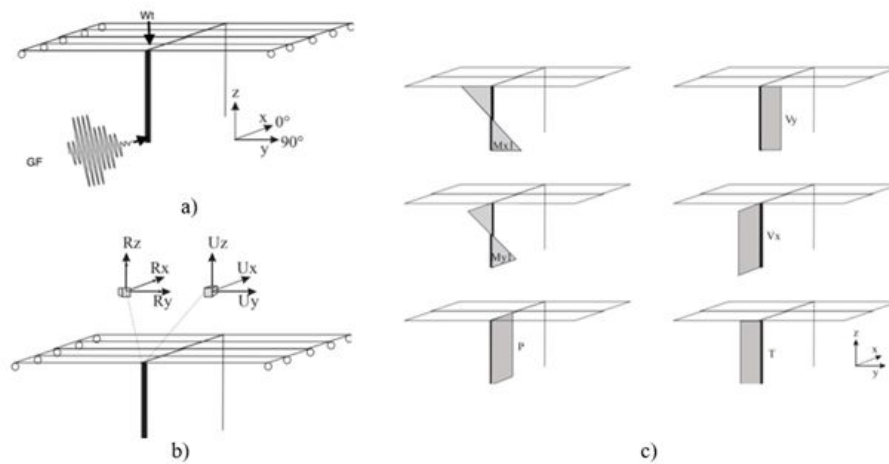


FIGURA 4. Variables utilizadas en la pila de un apoyo del viaducto elevado del tren México-Toluca. a) Variables del sismo, peso total del tren y de los pasajeros, b) Variables para los desplazamientos en el nodo superior y c) Variables utilizadas para los esfuerzos.

Preparación de los datos

En esta etapa se buscó mejorar la calidad del conjunto de datos, para ello, con la función *duplicated* en Python, se identificó que el 17.55 % de los registros del conjunto de datos del sismo de Chile eran duplicados, por lo que se eliminaron. Para detectar los outliers se utilizaron gráficas de caja para describir visualmente la distribución de datos [16] y posteriormente se reemplazaron los valores atípicos con los obtenidos del cálculo de los máximos o mínimos del rango intercuartil del conjunto de datos (diferencia entre el tercer y el primer cuartil de la columna analizada). Como resultado, se obtuvo la distribución

mostrada en la Tabla 1. Cada conjunto de datos está compuesto por 4 clases correspondientes a los factores de amplificación del sismo.

Tabla 1. Característica de los conjuntos de datos utilizados.

Conjunto de datos	Atributos	Clase 1	Clase 2	Clase 3	Clase 4
Sismo Chile	32	173	173	272	202
Sismo México	16	2314	2674	2557	2454

Modelado

Los modelos construidos se basaron en varios de los algoritmos más utilizados en el aprendizaje automático utilizando el software Weka en su versión 3.8.6, considerando los valores por default:

1. EM: maxIterations: 500, minStdDev: 1.0E-6, numClusters: 4, numExecutionSlots: 1, numFolds:10, numKMeansRuns: 10, seed: 100.
2. K-means: canopyMaxNumCanopiesToHoldInMemory:100, numClusters: 4, numExecutionSlots: 1, maxIterations:500, seed:10, canopyMinimumCanopyDensity: 2.0.
3. Multilayer Perceptron: learningRate: 0.3, momentum: 0.2, numDecimalPlaces:2, seed: 0, training Time: 500, validationSetSize: 0, validationThreshold: 20.
4. Arbol J48: confidenceFactor: 0.25, minNumObj: 2, numDecimalPlaces: 2, numFolds: 3, seed: 1.
5. SVM: c: 1, epsilon: 1.0E-12, numDecimalPlaces: 2, numFolds: -1, randomSeed: 1, toleranceParameter: 0.001.

Evaluación

En los experimentos se usó validación cruzada con 10 particiones, empleando el 80 % de los datos para entrenamiento y el 20 % para validación. Se evaluó el rendimiento de los modelos mediante dos medidas comunes en ciencia de datos: precisión general y media geométrica, derivadas de una matriz de confusión como la de la Tabla 2.

Tabla 2. Muestra una matriz de 2x2 para un problema de dos clases.

	Predicción Clase (+)	Predicción Clase (-)
Clase real (+)	Verdadero positivo (VP)	Falso negativo (FN)
Clase real (-)	Falso positivo (FP)	Verdadero negativo (VN)

En Tabla 2, VP y VN representan el número de instancias clasificadas por el modelo, mientras que FP y FN corresponden a errores cometidos de ambas clases. De este modo, la precisión general se expresa como Ec. (3):

$$Pg = \frac{VP + VN}{VP + FN + VN + FP} \quad (3)$$

Debido al desequilibrio existente en los datos, la utilización de la media geométrica Ec. (4), es ampliamente utilizada:

$$g = \sqrt{a^+ * b^-} \quad (4)$$

Donde: a^+ es la precisión de la clase positiva = $\frac{VP}{VP+FN}$ y a^- es la precisión de la clase negativa = $\frac{VN}{VN+FP}$

EXPERIMENTACIÓN

En el objetivo de esta investigación, se persiguen dos resultados: validar la distribución de instancias en la generación inicial de los conjuntos de datos y estimar el comportamiento de la estructura ferroviaria del tren interurbano si se presenta algún sismo de magnitud similar al del caso de estudio.

Agrupamiento

Para validar la distribución de datos en el conjunto creado con simulaciones en SAP 2000, se aplicaron los algoritmos EM y K -medias. La Tabla 3 muestra la distribución de instancias tras el agrupamiento del conjunto de datos de Chile. Las filas representan los valores de referencia y las columnas la asignación del agrupamiento.

Del agrupamiento obtenido se pueden hacer varios comentarios. La integración de la clase 3 en ambos agrupamientos es consistente y libre de ruido, como se observa en la intersección de la clase 3 con el grupo 3, donde los valores coinciden con los de referencia (Tabla 1). Para K -medias, la clase 2 prácticamente desaparece, asignándose a la clase 1. Esto se evidencia en las intersecciones entre las clases 2 y el grupo 2, así como entre la clase 1 y el grupo 1. Al momento de analizar el agrupamiento

obtenido con el conjunto de datos de sismo en México, no es posible encontrar una correspondencia lógica entre los grupos identificados y las clases existentes (Tabla 4). En todos los grupos obtenidos se encuentran instancias de las cuatro clases, lo que indica un problema en la calidad de los datos fuente.

Tabla 3. Distribución del agrupamiento del sismo en Chile.

	Grupo 1		Grupo 2		Grupo 3		Grupo 4	
	EM	k-medias	EM	k-medias	EM	k-medias	EM	k-medias
Clase 1	105	173	70	0	0	0	0	0
Clase 2	68	173	103	0	0	0	0	0
Clase 3	0	0	0	0	272	272	0	0
Clase 4	0	0	0	89	0	0	202	113
Total	173	347	173	89	272	272	202	113

Tabla 4. Distribución del agrupamiento del sismo en México.

	Grupo 1		Grupo 2		Grupo 3		Grupo 4	
	EM	k-medias	EM	k-medias	EM	k-medias	EM	k-medias
Clase 1	1241	126	1219	227	996	274	608	358
Clase 2	297	1863	562	1789	636	1343	841	758
Clase 3	639	188	580	345	447	462	283	616
Clase 4	137	137	313	313	478	478	722	722
Total	2314		2674		2557		2454	

Clasificación

La Tabla 5 presenta el promedio de precisión general y la media geométrica, destacando en negrita los mejores resultados. Utilizar las distribuciones obtenidas con los algoritmos de agrupamiento mejora sustancialmente la clasificación, con poca variación entre algoritmos de clasificación. En su mayoría, los mejores resultados son obtenidos con el algoritmo EM.

Tabla 5. Resultados de la clasificación.

CD	Medida	PM		SVM		J48	
		Chile	México	Chile	México	Chile	México
Original	Pg	80.8	40.68	76.9	40.74	80.2	37.33
	g	85.1	57.2	82	57.7	86.6	54.6
EM	Pg	97.8	99.99	97.6	99.99	98.3	99.99
	g	98.3	99.9	98.1	99.9	98.7	99.9
k-medias	Pg	99.3	99.5	99	99.62	98.7	99.12
	g	99.1	99.5	98.8	99.6	98.5	99.1

La Tabla 6 muestran en promedio la matriz de confusión obtenida al utilizar los modelos de ML con el conjunto de datos original, es decir, sin agrupar.

Tabla 6. Distribución del agrupamiento del sismo en México.

Clase	Sismo Chile				Sismo México			
	Perceptrón Multicapa		Support Vector Machine		Clasificador J48			
	Chile	México	Chile	México	Chile	México	Chile	México
Perceptrón Multicapa								
Clase 1	242	22.33	0	0	2027.7	299.3	72.7	54.7
Clase 2	30	162.3	0	2	203.7	3098.7	128.7	157.7
Clase 3	0	0.33	194.7	0.7	109.3	357.3	1319	253.3
Clase 4	0	4	1.3	161.3	31	206.7	120	1560.3
Support Vector Machine								
Clase 1	232.7	31.7	0	0	2159.7	214.7	46.3	33.7
Clase 2	31.7	159	0	3.7	351.7	3006.7	96	134.3
Clase 3	0	0	194.7	1	210.7	297.7	1295.7	235
Clase 4	0	3	1.7	162	72.7	180.7	115	1549.7
Clasificador J48								
Clase 1	256	8.3	0	0	2021	225.7	145.3	62.3
Clase 2	46	145.7	0.3	2.33	219.3	2962.7	239.3	167.3
Clase 3	0	0	194	1.67	148.3	250.7	1399.3	240.7

Tabla 6. Distribución del agrupamiento del sismo en México (...continuada).

Clase	Sismo Chile				Sismo México			
	Perceptrón Multicapa		Support Vector Machine		Clasificador J48			
	Chile	México	Chile	México	Chile	México	Chile	México
Clase 4	0.3	2	1.7	162.7	54	152	213.7	1498.3

Las matrices de confusión anteriores destacan el traslape de clases en el conjunto de datos del sismo de México, lo que justifica la baja precisión observada en los tres clasificadores utilizados.

Una estrategia estadística común para analizar la concordancia en los resultados es el análisis del coeficiente Kappa [17], que compara la concordancia observada en un conjunto de datos, respecto a la que podría ocurrir por mero azar. Estos resultados se muestran en la Tabla 7.

Tabla 7. Coeficiente Kappa obtenido con los diferentes modelos utilizados.

CD	Sismo	PM	SVM	J48
Original	Chile	0.7401	0.6874	0.7317
	México	0.2066	0.2115	0.1637
EM	Chile	0.9703	0.9669	0.9768
	México	0.9999	0.9999	0.9999
K-medias	Chile	0.9893	0.9857	0.9803
	México	0.9918	0.9937	0.9855

Los valores de kappa varían de -1 a $+1$, mientras más alto sea el valor, más fuerte es la concordancia. De la Tabla 7, es posible ver que los resultados obtenidos con el conjunto de datos del sismo de Chile cuentan con un grado de acuerdo, reforzando las conclusiones obtenidas.

CONCLUSIONES

Los resultados muestran que el análisis de clúster revela un severo traslape de clases en los datos sintéticos originales del sismo en México comparado con Chile. Esto se debe a la dificultad para encontrar una correspondencia lógica entre los grupos identificados y las clases en los datos de México. La distribución obtenida por los algoritmos de agrupamiento cuestiona la distribución generada inicialmente con el software SAP 2000 y se espera que afecte el rendimiento de los modelos al intentar aprender a clasificar.

Por otro lado, en lo relacionado a la clasificación considerando tanto las clases inicialmente definidas, así como las identificadas con los algoritmos de agrupamiento, fue posible observar que, al ser entrenados con los conjuntos derivados del agrupamiento, mejora su desempeño, principalmente con el algoritmo EM.

Las líneas de estudio abiertas se dirigen hacia dos aspectos: aplicar técnicas más profundas de limpieza de datos para resolver el problema de traslape de clases antes de la clasificación y utilizar métodos para generar datos sintéticos que conserven las distribuciones estadísticas, aumentando el volumen de datos y mejorando el rendimiento de los clasificadores.

Es importante tener en cuenta que las circunstancias sobre las que se hizo el estudio de la estructura fueron muy exigentes, y las probabilidades de que se produzcan las cargas que se analizaron son escasas. La estructura está dentro de los parámetros de seguridad establecidos, aunque se requiere más investigaciones en futuros trabajos para identificar las condiciones críticas de seguridad.

Finalmente, es necesario comentar que este artículo forma parte de un proyecto de investigación (UAEMex 6523/2022 CIA) que pretende establecer la probabilidad de falla de un “tramo tipo” del viaducto elevado del tren México-Toluca, sujeto simultáneamente a sismo y carga viva máxima. Publicaciones posteriores complementarán, entonces, la hasta aquí presentada.

REFERENCIAS

- [1] Secretaría de Obras y Servicios de la CDMX. *Tren El Insurgente*. Recuperado el 15 de junio de 2024, de <https://www.obras.cdmx.gob.mx/proyectos/tren-interurbano-cdmx>
- [2] H.-T. Thai. Machine learning for structural engineering: A state-of-the-art review. *Structures*, vol. 38, pp. 448–491, Apr. 2022. DOI: 10.1016/j.istruc.2022.02.003
- [3] S. Mangalathu and J.-S. Jeon. Machine Learning-Based Failure Mode Recognition of Circular Reinforced Concrete Bridge Columns: Comparative Study. *Journal of Structural Engineering*, vol. 145, pp. 04019104, Oct. 2019. DOI: 10.1061/(ASCE)ST.1943-541X.0002402
- [4] D. H. Nguyen, Q. B. Nguyen, T. Bui-Tien, G. De Roeck, and M. Abdel Wahab. Damage detection in girder bridges using modal curvatures gapped smoothing method and Convolutional Neural Network: Application to Bo Nghi bridge. *Theoretical and Applied Fracture Mechanics*, vol. 109, pp. 102728, Oct. 2020. DOI: 10.1016/j.tafmec.2020.102728

- [5] H. Mahmoudi, M. Bitaraf, M. Salkhordeh, and S. Soroushian. A rapid machine learning-based damage detection algorithm for identifying the extent of damage in concrete shear-wall buildings. *Structures*, vol. 47, pp. 482–499, Nov. 2022. DOI: 10.1016/j.istruc.2022.11.041
- [6] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, vol. 388, pp. 154–170, Feb. 2017. DOI: 10.1016/j.jsv.2016.10.043
- [7] M. Flah, I. Nunez, W. Ben Chaabene, and M. L. Nehdi. Machine Learning Algorithms in Civil Structural Health Monitoring: A Systematic Review. *Archives of Computational Methods in Engineering*, vol. 28, no. 4, pp. 2621–2643, Jun. 2021. DOI: 10.1007/s11831-020-09471-9
- [8] S. Sarothi, K. Ahmed, N. Imtiaz, A. Ahmed, and M. Nehdi. Machine Learning-Based Failure Mode Identification of Double Shear Bolted Connections in Structural Steel. *Engineering Failure Analysis*, vol. 139, pp. 106471, 2022. DOI: 10.1016/j.engfailanal.2022.106471
- [9] M. Abdellatif, H. Peel, A. G. Cohn, and R. Fuentes. Combining block-based and pixel-based approaches to improve crack detection and localization. *Automation in Construction*, vol. 122, pp. 103492, Feb. 2021. DOI: 10.1016/j.autcon.2020.103492
- [10] A. H. Alkadhwi and O. Adelaja. Data mining application using clustering techniques (K-Means algorithm) in the analysis of student's result. *Journal of Multidisciplinary Engineering Science Studies (JMESS)*, vol. 5, no. 5, pp. 2458–925, 2019. ISSN: 2458-925X.
- [11] B. Pani, J. Klemenc, and M. Nagode. Improved Initialization of the EM Algorithm for Mixture Model Parameter Estimation. *Mathematics*, vol. 8, no. 3, p. 373, 2020. DOI: 10.3390/math8030373
- [12] D. A. Alboaneen, H. Tianfield, and Y. Zhang. Sentiment analysis via multi-layer perceptron trained by meta-heuristic optimisation. In *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 4630–4635. DOI: 10.1109/BigData.2017.8258507
- [13] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020. DOI: 10.1016/j.neucom.2019.10.118
- [14] J. Jiménez. Función del enlace técnico dentro de la obra del tren Interurbano México Toluca. *Memoria de Experiencia Laboral*, Ingeniería Civil, Facultad de Ingeniería, Universidad Autónoma del Estado de México, 2017.

- [15] D. J. Delgado Hernández, R. A. Mejía Nava, D. De León Escobedo, J. C. Arteaga Arcos, M. Á. Mendoza Lugo, and J. F. García Zamora. Análisis de riesgos del tren México-Toluca sujeto simultáneamente a sismo y carga viva máxima. Informe del proyecto 6523/2022CIA. SIEA, Universidad Autónoma del Estado de México, 2023.
- [16] R. Arora and S. Suman. Comparative Analysis of Classification Algorithms on Different Datasets using WEKA. *International Journal of Computer Applications*, vol. 54, pp. 21–25, 2012.
- [17] J. R. Dettori and D. C. Norvell. Kappa and Beyond: Is There Agreement? *Global Spine Journal*, vol. 10, no. 4, pp. 499–501, 2020. DOI: 10.1177/2192568220911648
- [18] K. Stapor, P. Ksieniewicz, S. García, and M. Woniak. How to design the fair experimental classifier evaluation. *Applied Soft Computing*, vol. 104, p. 107219, Jun. 2021. DOI: 10.1016/j.asoc.2021.107219
- [19] J. B. Rollins. Metodología Fundamental para la Ciencia de Datos. *IBM, Armonk, NY, Technology, IMWI4828-COES-01*, 2015.