

# Predicción de Expansión Urbana de las Ciudades Zacatecas-Guadalupe Usando Maquinas de Soporte Vectorial

A. Carmina Llamas-Valenzuela <sup>[0009-0003-4646-1830]</sup>, José I. de la Rosa <sup>[0000-0002-7337-8974]</sup>,  
G. Moreno-Chávez <sup>[0000-0002-2498-138X]</sup>, Efrén Gonzales-Ramírez <sup>[0002-0002-8060-6170]</sup>,  
Jesús Villa <sup>[0000-0002-5395-855X]</sup>, José M. Celaya-Padilla <sup>[0000-0001-6847-3777]</sup>

Universidad Autónoma de Zacatecas,  
Unidad Académica de Ingeniería Eléctrica,  
México

ismaelrv@ieee.org, {allamas, gamalielmch,  
gonzalezefren, jvillah, jose.celaya} @uaz.edu.mx

**Resumen.** Las ciudades son espacios físicos en los cuales se establece la población. En México, tres de cada cuatro personas viven en una ciudad. La expansión rápida y sin planeación trae consigo consecuencias indeseables para el desarrollo social y económico. El objetivo de este estudio es modelar y predecir la expansión urbana en la zona metropolitana Zacatecas-Guadalupe usando máquinas de soporte vectorial, y así realizar una mejor planeación. Para lograr este objetivo se utilizaron mapas de uso de suelo y cobertura terrestre correspondientes al periodo 2000 a 2020, así como la inclusión de variables socioeconómicas, topográficas y atributos culturales. Se desarrolló un modelo de SVM penalizado con una exactitud de entrenamiento de un 92.4%, una exactitud de validación 93 % y un F1-Score de 86.3 %. En los resultados obtenidos, se puede observar que la cercanía hacia áreas ya urbanizadas y el tipo de uso de suelo tienen una alta influencia en la urbanización. Adicionalmente, la pendiente del terreno tiene muy poca influencia en las decisiones de urbanización. Comparado con otros estudios, este incorpora variables culturales e integra los valores SHAP, con el objetivo de conocer la influencia de dichas variables en el modelo final.

**Palabras clave:** expansión urbana, máquinas de soporte vectorial, predicción.

## A Prediction of Urban Expansion of the Cities of Zacatecas-Guadalupe Using Support Vector Machine

**Abstract.** Cities are physical spaces in which the population is established. In Mexico, three out of four people live in a city. Rapid and unplanned expansion brings with it undesirable consequences for social and economic development. The objective of this study is to model and predict urban expansion in the

Zacatecas-Guadalupe metropolitan area using support vector machines, and thus carry out better planning. To achieve this objective, land use and land cover maps corresponding to the period 2000 to 2020 were used, as well as the inclusion of socioeconomic, topographic and cultural attributes variables. A penalized SVM model was developed with a training accuracy of 92.4%, a validation accuracy of 93% and an F1-Score of 86.3%. In the results obtained, it can be observed that the proximity to already urbanized areas and the type of land use have a high influence on urbanization. Additionally, the slope of the land has very little influence on urbanization decisions. Compared to other studies, this one incorporates cultural variables and integrates SHAP values, with the objective of knowing the influence of these variables on the final model.

**Keywords:** urban expansion, support vector machine, prediction.

## 1 Introducción

Las ciudades son espacios físicos en los cuales se establece la población. En México, tres de cada cuatro personas viven en una ciudad [1]. Este incremento ocasiona el crecimiento urbano. El crecimiento es un proceso socioeconómico complejo que involucra la transformación de espacios destinados a actividades primarias a actividades terciarias [2].

El crecimiento rápido y sin planeación trae consigo consecuencias indeseables para el desarrollo social y económico. Algunas de las consecuencias son: el daño ecológico [3], insuficientes viviendas para la demanda, sistemas de transporte deficientes, aumento en la segregación social [4] y el incremento del tráfico [5]. Estas consecuencias se pueden aminorar, realizando una planeación a largo plazo, apoyada de diferentes instrumentos [6].

A lo largo de los años, expertos en el área de planeación urbana han desarrollado diferentes métodos para el modelado de la expansión urbana. Algunos de estos modelos se apoyan de datos obtenidos de la percepción remota, en los cuales se utilizan imágenes satelitales. Estas se emplean para conocer el uso de suelo y cobertura de la tierra. Adicionalmente, estos modelos se apoyan ampliamente en el uso de Sistemas de Información Geográfica (GIS, por sus siglas en inglés) [7]. Es importante aclarar que estos esfuerzos de prever la expansión urbana no sustituyen el conocimiento de los desarrolladores inmobiliarios o académicos en la planeación de ciudades. Los instrumentos deben considerarse como un apoyo a la toma de decisiones.

Los modelos matemáticos utilizados inicialmente incluyen los basados en autómatas celulares [4, 8–10] y modelos basados en agentes (ABM) [11]. Estos han sido ampliamente utilizados en diferentes estudios. Sin embargo, estos modelos asumen que las áreas urbanas son espacialmente homogéneas. Por esta razón, el modelado de las decisiones individuales e interacciones socioeconómicas es difícil de representar [12].

Con la intención de solucionar esta falta de representación, se comenzó a incorporar el aprendizaje automático, como: XGboost-SHAP [5], máquinas de soporte (SVM) [13, 14], árboles de decisión [15], bosques aleatorios [16], entre otros. Sin embargo, estos estudios se realizaron fuera de México. El hecho de que para el 2020 casi cien millones

de mexicanos vivían en asentamiento urbanos no planificados [4] impulsaron la creación de este estudio.

El objetivo de este estudio es modelar y predecir la expansión urbana en la ciudad Zacatecas-Guadalupe usando máquinas de soporte vectorial. Para lograr este objetivo se utilizaron mapas de uso de suelo y cobertura terrestre correspondientes al periodo 2000 a 2020, así como la inclusión de variables socioeconómicas, topográficas y atributos culturales. En las siguientes secciones se describen los materiales y métodos usados en el estudio (sección 2). Posteriormente, se analizan y discute lo encontrado (sección 3). Finalmente, se concluye con recomendaciones para estudios futuros (sección 4).

## **2 Materiales y métodos**

### **2.1 Área de estudio**

El estado de Zacatecas está ubicado en la zona centro-norte del país. Convirtiéndolo en un estado de alta conectividad por vía terrestre. La razón de esta situación radica en la proximidad de las ciudades de Aguascalientes, San Luis Potosí y Durango [17].

Lo anterior contribuyó a la creación de la zona metropolitana Zacatecas-Guadalupe, conformada por cinco municipios. De acuerdo con [18] la metrópolis Zacatecas-Guadalupe tuvo una tasa de crecimiento medio anual de 2.03 % en el período de 2010-2020. Este trabajo está enfocado en los municipios centrales: Zacatecas y Guadalupe.

### **2.2 Datos**

Con el fin de examinar y ejecutar la predicción de la zona Zacatecas-Guadalupe, se elaboró un conjunto de datos con veinte variables compuesto por variables topográficas, socioeconómicas, ambientales, culturales y de vecindario de los años 2000 a 2020 (Tabla 1). Las variables se seleccionaron tomando en cuenta estudio relacionados [4, 5, 13, 15, 19, 20], estadísticas religiosas del estado [21] y otras características poblacionales.

La construcción de las variables se llevó a cabo utilizando datos de los censos del Instituto Nacional de Estadística y Geografía (INEGI) y otras fuentes. Fueron preprocesadas con el software de uso libre QGIS. Debido a que los Mapas de Uso de Suelo se obtuvieron de imágenes multiespectrales de resolución espacial de 30 m x 30 m, todas las variables se ajustaron a esa resolución y se normalizaron.

En la categoría de topografía se encuentran variables que describen el terreno. Se usó el Modelo Digital de Elevación con una resolución de 30 m por celda [22]. A partir del Modelo Digital de Elevación, se calculó la pendiente y la pendiente usando QGIS,

Las variables de distancias, son distancias euclidianas calculadas en metros. Se calcularon a partir de los archivos vectoriales publicados por INEGI. Por último, se calcularon las variables correspondientes a densidad de población, viviendas e ingresos a partir de los censos correspondientes a los años 2000 [23], 2010 [24] y 2020 [25], realizados por INEGI.

**Tabla 1.** Variables utilizadas en el estudio.

<b>Categoría</b>	<b>Variable</b>
Topografía	(1) Modelo Digital de Elevaciones, (2) pendiente del terreno y (3) dirección de la pendiente
Uso de Suelo	(4) Mapas de uso de suelo, (5) Distancia a áreas construidas, (6) calles y carreteras, (7) vías férreas y (8) centro de las ciudades.
Ambientales	(9) Distancia a áreas verdes/parques
Socioeconómicas	(10) Densidad de población, (11) población ocupada, (12) jefatura femenina y (13) jefatura masculina, (14) ingresos trimestrales por vivienda y (15) número de viviendas por celda.
Culturales	(16) Distancia a templos, (17) cementerios y (18) escuelas, (19) densidad de población religiosa,
Vecindario	(20) Probabilidad de cambio del pixel

### 2.3 Metodología

En la Figura 1 se presenta la metodología general la cual se divide en dos etapas: (1) desarrollar el modelo de expansión urbano utilizando los datos de 2000 y 2010 y (2) la predicción de la expansión urbana para la zona Zacatecas-Guadalupe, usando los datos del año 2020.

En la etapa uno se llevó a cabo la recolección de datos y su preprocesamiento. Se seleccionaron y limpiaron los datos relevantes, eliminando valores atípicos y faltantes. Luego, se transformaron en imágenes para su análisis posterior. En este estudio se seleccionaron los años 2000, 2010 y 2020. Las variables fueron preprocesadas usando QGIS o Python.

Teniendo los datos preprocesados, se realizó el modelado usando el algoritmo de Máquinas de Soporte Vectorial (SVM). La selección del algoritmo se debió a que es robusto y ha sido utilizado en estudios relacionado con buenos resultados [13]. Los datos se dividieron como indica la Tabla 2. La variable dependiente está descrita como -1 para píxeles no construidos y 1 para píxeles construidos. En la Figura 2 se muestran los datos de entrenamiento. Debido a la disparidad de las clases, se realizó un submuestreo aleatorio nivelando las clases.

**Tabla 2.** División de datos en entrenamiento y validación.

<b>Conjunto de datos</b>	<b>Variables independientes</b>	<b>Variable dependiente</b>
Entrenamiento	Datos del año 2000	Clasificación binaria del 2010
Validación	Datos del año 2010	Clasificación binaria del 2020

En la etapa final (2), se realizó el cálculo de los valores SHAP y la predicción de la zona Zacatecas-Guadalupe al año 2030, usando el mejor modelo. Para la predicción se usaron los datos topográficos, socioeconómicos, ambientales, culturales y de vecindario del año 2020. El resultado es una imagen de valores -1 (no construido) y 1 (construidos), de resolución espacial de 30 m x 30 m.

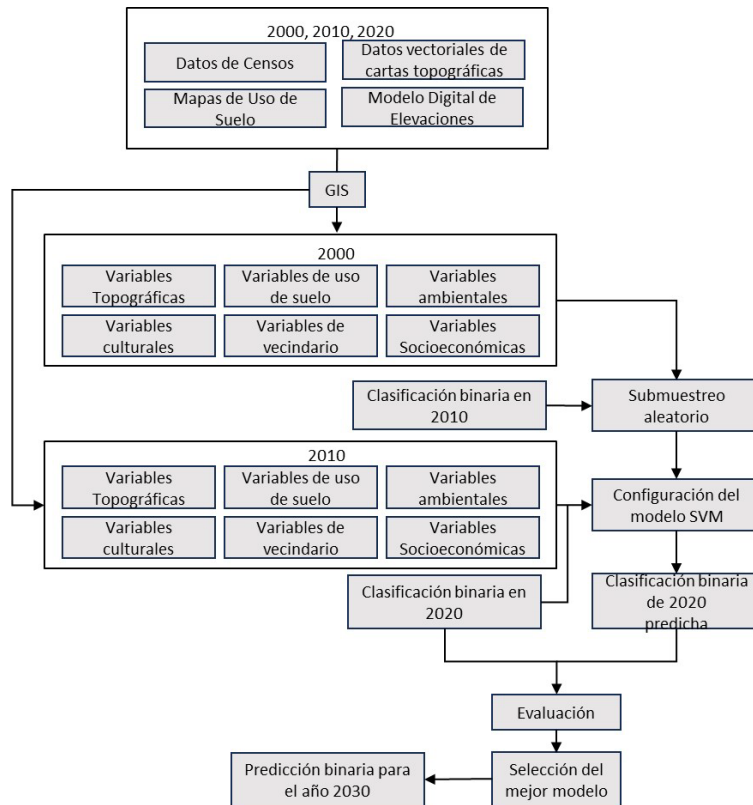


Figura 1. Diagrama de la metodología seguida en este estudio.

#### 2.4 Máquinas de soporte Vectorial (SVM) con penalización.

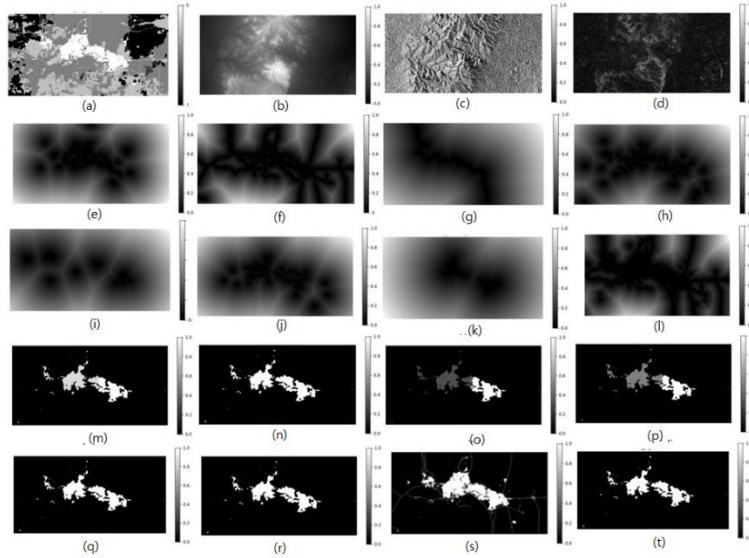
Las máquinas de soporte vectorial fueron propuestas por Vapnik and Lerner (1963). Se introdujeron inicialmente para la clasificación binaria supervisada [26]. Esta técnica se fundamenta en la construcción de un Hiperplano de Separación Óptimo (HSO) y la construcción de los vectores de soporte.

Supongamos que tenemos un conjunto de entrenamiento  $\{(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_N, y_N)\}$ , donde,  $(\vec{x}_i, y_i)$  corresponden a las muestras para  $i=1 \dots N$  está constituido por un vector  $n$  de características y una etiqueta que indica las clases  $\{\pm 1\}$  a la que pertenece a cada una de las muestras [26].

Debido a que las clases son +1 y -1, el HSO está definido por el margen máximo de separación entre las clases. Con base en esto, los vectores de soporte se definen como:  $\vec{w} \cdot \vec{x} + b = +1$  y  $\vec{w} \cdot \vec{x} + b = -1$ , los cuales son paralelos al HSO el cual se expresa por  $\vec{w} \cdot \vec{x} + b = 0$ . El margen máximo de los vectores de soporte está expresado por  $2/\|\vec{w}\|$ .

Sin embargo, inicialmente las SVM fueron propuestas de manera rígida, es decir, que los datos son completamente separables. No obstante, existen datos que no lo son.

En estos casos se introduce una variable de holgura  $\xi_i$ , la cual permite errores en la clasificación, pero estos errores son penalizados.



**Figura 2.** Datos de entrenamiento del año 2000 usadas en el estudio. (a) Uso de Suelo, (b) Modelo Digital de elevaciones, (c) Dirección de la pendiente, (d) Pendiente. (e)-(l) variables de distancia y (m)-(t) Variables de densidad.

Al incluir este nuevo parámetro la SVM se suaviza. El problema primario cambia de forma y ahora se busca minimizar la expresión:

$$J(w, \xi) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i,$$

$$\text{Sujeta a la condición: } y_i [w_i \cdot x_i + b] + \xi_i - 1 \geq 0, \xi_i \geq 1, i = 1, \dots, n \quad (1)$$

Donde  $\xi_i$  es el parámetro de holgura y  $c$  es el parámetro de penalización a los errores de clasificación [13]. En este nuevo enfoque, el objetivo es encontrar el HSO que minimice los errores de clasificación y maximice la separación entre los vectores de soporte. Mientras el valor de  $c$  se incrementa, se obtiene un margen estrecho, esto minimiza el número de clasificaciones erróneas. Por otro lado, si  $c$  disminuye, se están permitiendo más clasificaciones erróneas [27].

Para abordar la no linealidad, los datos son transportados a un espacio de alta dimensionalidad. Esto se logra realizando una proyección matemática de los datos usando un kernel. Dicho kernel se puede definir como:  $k(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$ . Algunas funciones de kernel populares son la lineal  $k(x, y) = x_i \cdot x_j$ , función de base radial (RBF)  $k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$ , y la función polinomial  $k(x, y) = (1 + x_i \cdot x_j)^q$ . Donde  $\gamma$  y  $q$  son parámetros correspondientes al kernel [13].

## 2.5 Métricas de evaluación

Al entrenar un modelo de aprendizaje automático, es necesario evaluar el rendimiento del modelo. Existen diferentes técnicas de evaluación; la elección depende del

problema y su contexto. En la clasificación, algunas de las evaluaciones disponibles son exactitud, precisión, área bajo la curva, entre otros [28]. Estas métricas se calculan usando la matriz de confusión [13], la cual nos introduce a los términos de verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN).

Como se puede visualizar en la Tabla 3, los TP se refiere a los píxeles que son construidos y el modelo los clasificó como construidos. FP indica los píxeles que son no construidos y el modelo lo colocó como construidos. TN, píxeles clasificados como no construidos y en la etiqueta real son no construidos. Por último, FN píxeles que por el modelo se indicaron como no construidos, pero en la etiqueta real son construidos[28].

En este estudio se utilizó: la exactitud, precisión y F1-Score. La exactitud indica la proporción de clasificaciones correctas realizadas por el modelo respecto al total de muestras, se calculó usando (2). La fórmula (3) se utiliza para calcular la precisión, la cual se centra en la proporción de TP respecto al total de muestras clasificadas como positivas. Por último, el F1-Score calculado con la ecuación (5) es la media armónica de la precisión y recall (ecuación 4). El recall se refiere a la proporción de TP respecto a las instancias que son positivas [29]:

$$\text{Exactitud} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

$$\text{Precisión} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F1-score} = \frac{2 * (\text{Recall} * \text{Precisión})}{(\text{Recall} + \text{Precisión})} \quad (5)$$

**Tabla 3.** Matriz de confusión de píxeles construidos-no construidos.

Predicción observada	Construido	No-construido
Construido	Verdaderos Positivos (TP)	Falsos Negativos (FN)
No construido	Falsos Positivos (FP)	Verdaderos Negativos (TN)

## 2.6 Shapley Additive exPlanations (SHAP)

SHAP fue propuesto por Lundberg y Lee [30], se utiliza para asignar un valor de importancia a cada característica para una predicción específica. Esto se basa en la teoría de los valores Shapley, que proviene de la teoría de juegos cooperativos. Los valores Shapley miden la contribución marginal de cada característica al resultado final utilizando la siguiente fórmula [5]:

$$\Phi_i = \sum_{S \in N} \frac{|S|! (n - |S| - 1)!}{n!} [v(S \cup \{i\}) - v(S)] \quad (6)$$

Donde  $N$  es el número de instancias que tienen  $n$  características,  $S$  representa un subconjunto de las características  $n$  y  $v$  representa las características de entrada del conjunto  $S$ .

### 3 Resultados

Antes de desarrollar el modelo predictivo se llevó a cabo un análisis de los píxeles correspondientes a las etiquetas de construido y no construido. Con esto se calculó la tasa de cambio de los píxeles entre los años estudiados (Tabla 4). La tasa de cambio se mantuvo constante en los dos periodos analizados. El aumento está enfocado principalmente al noreste en el municipio de Guadalupe. En el municipio de Zacatecas, el crecimiento está focalizado en el área suroeste de la ciudad.

**Tabla 4.** Tasa de crecimiento de la zona de estudio en los periodos 2000 a 2010 y 2010 a 2020.

	2000-2010	2010-2020
Cantidad de píxeles que cambiaron	16369	14578
Cantidad de píxeles que no cambiaron	343133	344924
Tasa de cambio	4.55%	4.06%

#### 3.1 Modelo SVM

Se realizó una experimentación modificando diferentes valores de los parámetros del SVM penalizado, basado en el trabajo [13]. Los parámetros que tomaron diferentes valores en esta experimentación fueron: parámetro  $c$ , kernel, gama y grado  $q$  del polinomio. Para el parámetro  $c$  se seleccionaron los valores 0.1, 1, 10, 100. Asimismo, se seleccionó el kernel radial (RBF) con el parámetro gama de 1, 2, 3. Finalmente, para el kernel polinomial se eligieron los valores de grado de polinomio de 1, 2, 3, 4 y 5. Las combinaciones se realizaron con el propósito de obtener el modelo predictivo más adecuado.

La selección del modelo se llevó a cabo usando las métricas de exactitud (entrenamiento y validación), F1-Score y Precisión. Estas métricas se calcularon usando la matriz de confusión de los píxeles construidos y no construidos.

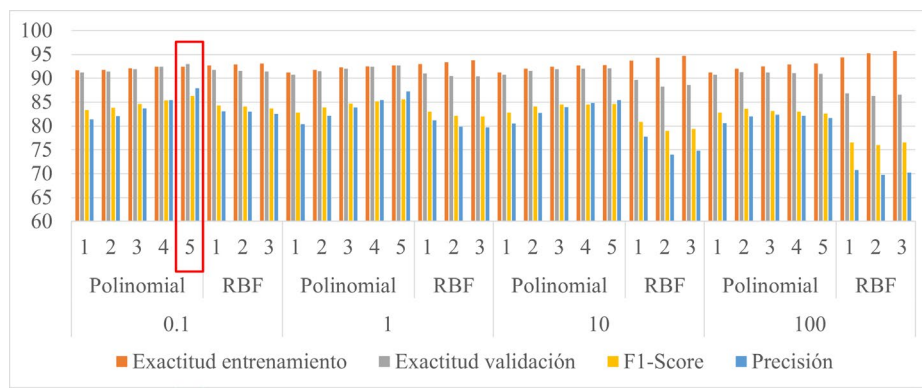
Los modelos resultantes tienen una exactitud de entrenamiento y validación elevada (mayores al 90 %). En la Figura 3 se pueden apreciar los resultados. En la exactitud de entrenamiento, los valores más altos son aquellos obtenidos con los modelos con kernel RBF. Por otro lado, en la exactitud de validación, los mejores valores se encuentran en los modelos con kernel polinomial con grado de polinomio de cinco. Sin embargo, debido al desbalance de las clases, se decidió apoyarnos del F1-Score.

Los resultados más elevados en el F1-Score se encuentran con el kernel polinomial y grado del polinomio cinco, similar a la exactitud de validación. Estos resultados están alrededor del 80 %. Los resultados en la precisión son similares al F1-Score, los mejores se encuentran con el kernel polinomial en los valores del 80 %.

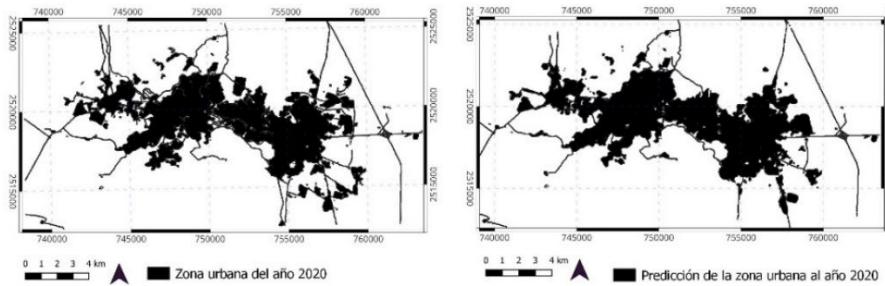


En el caso de los valores de  $c$  los mejores resultados se encuentran en los valores de 100 y 0.1. En la exactitud presentada en el entrenamiento y precisión, los mejores valores están con el valor  $c$  de 100. Mientras que la exactitud de validación y F1-Score se encuentran con el valor de 0.1.

El modelo que se eligió como el mejor, se debe a que tiene uno de los mejores F1-Score y un balance con las demás métricas. Este modelo es el de kernel polinomial, con valor  $c$  de 0.1 y un grado de polinomio de cinco, la exactitud de entrenamiento es de un 92.4%, exactitud de validación 93% y F1-Score de un 86.3%. En la Figura 3 en el recuadro rojo se indica el modelo seleccionado. En la Figura 4 se encuentra la comparación del resultado del modelo respecto a la realidad del año 2020.



**Figura 3.** Resultado de la experimentación, en el cuadro rojo se encuentra el que se considera el mejor modelo.



**Figura 4.** A la izquierda se encuentra las zonas urbanizadas en el año 2020, en la de la derecha la predicción de urbanización realizada por el modelo SVM penalizado.

### 3.2 Factores de importancia

Los valores SHAP se calcularon seleccionando un subgrupo de los datos. La selección se llevó a cabo aplicando el algoritmo de K-means a los datos y posteriormente seleccionando los puntos más cercanos a los centroides. En la Figura 5 se observan los valores SHAP, lo cual nos ayuda a comprender el comportamiento de las variables independientes que determinan si los píxeles se urbanizan o no.

La variable que más impacta al modelo es el tipo de uso de suelo. Mientras más alto es el valor, es más probable que se urbanice. Esto quiere decir que si ya es urbanizado permanece urbanizado y si es suelo desnudo o vegetación cercana a zona urbana, el clasificador lo coloca como urbanizado.

La segunda variable que tiene mayor influencia es la probabilidad de urbanización. Esto indica que el pixel se urbaniza en dependencia de cuantos píxeles dentro de su vecindario de 3x3 ya es urbanizado. Si el valor de esta probabilidad es alto el modelo lo indica como urbanizado. Esto va de la mano con la tercera variable la cual es la distancia a zona urbana. Si la distancia a la zona urbana es pequeña el modelo tiende a colocarlo como urbanizado. Por otro lado, la distancia a los centros de la ciudad es la cuarta variable que tiene importancia. Cuanto menor la distancia al centro, es más llamativa a la urbanización.

Por el contrario, las variables que no tuvieron un impacto fueron la elevación del terreno, los ingresos trimestrales, la pendiente y la dirección.

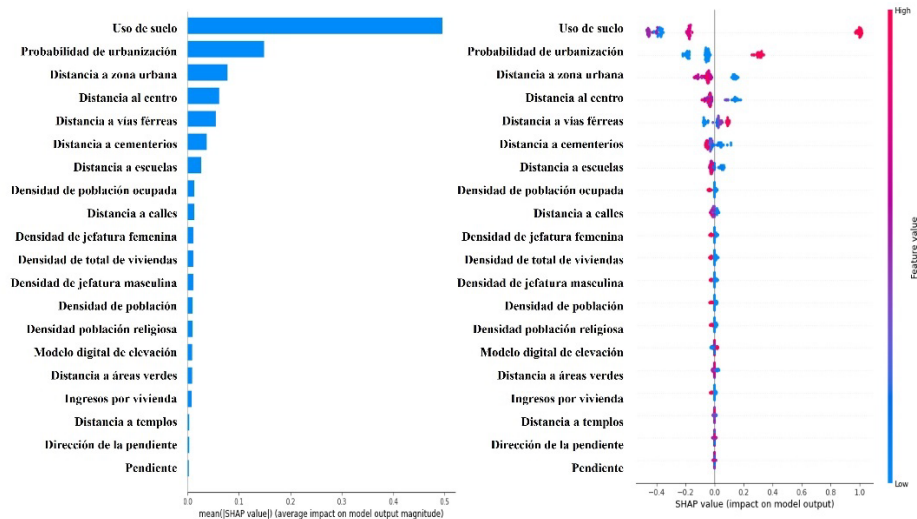
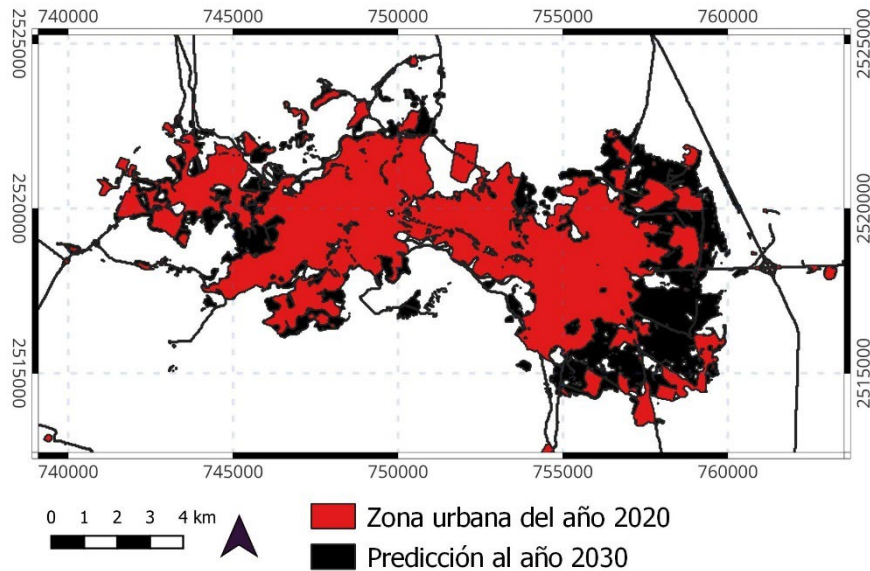


Figura 5. Valores SHAP de las diferentes variables del estudio.

### 3.3 Predicción del crecimiento urbano

Con el modelo de SVM penalizado desarrollado y los datos del año 2020 se realizó la predicción de las zonas urbanas para el año 2030, el cual se muestra en la Figura 6. En el resultado se puede observar que el crecimiento de la zona metropolitana es mayor en la zona de Guadalupe. Esto es algo que ha estado constante a lo largo de los años.

En cuanto a la dirección de la expansión en la zona de Guadalupe es más densa en la zona sureste de la zona de estudio. De forma contraria la expansión en el municipio de Zacatecas esta más concentrada en la zona central-oeste. Disminuyendo la expansión de la zona suroeste de los años anteriores. La tasa de cambio de los píxeles en esta predicción es del 8.11%. El doble de lo encontrado en el análisis de los periodos anteriores.



**Figura 6.** Zona urbanizada en el año 2020 y la predicción al año 2030 resultado del modelo predictivo.

#### 4 Discusión

La rápida urbanización es la causa principal del cambio de uso de suelo de áreas vegetales, agricultura y suelo desnudo a urbanización [31]. En este estudio, hemos explorado los patrones y tendencias del crecimiento urbano en la ciudad de Zacatecas-Guadalupe durante el periodo de 2000 a 2020. Utilizando SVM penalizado, el cual es un algoritmo robusto, e integrando los valores SHAP.

Como resultado de los valores SHAP se pudo observar que la cercanía a áreas ya urbanizadas y si el píxel es principalmente suelo desnudo, existe más probabilidad de que este se urbanice. Lo cual es consistente con otros estudios como [13, 15]. Sin embargo, a pesar de que algunos autores consideran importante la elevación y pendiente [5], en los valores SHAP encontrados en este estudio indican que estas variables no tienen impacto en la urbanización de la zona, lo cual es algo que encuentran otros autores [31, 32].

Las distancias a los centros de la ciudad, a escuelas y a calles también tienen un impacto en la predicción final. Lo cual también se ha visto en otros artículos [13]. En la predicción del 2030 se observó que la distancia a calles tiene una alta influencia. Ya que la predicción está concentrada en zonas que tienen más vías principales de primer y segundo orden, ubicadas en el área de Guadalupe. Mientras que en el área de Zacatecas se tiene una menor densidad de este tipo de vías, lo que provoca menor expansión en esa dirección.

Hablando de las variables culturales, distancia a templos y densidad de población religiosa no tuvieron un impacto significativo de acuerdo a los resultados. Por lo tanto, se consideraría utilizar otras variables que sean más llamativas para habitar un lugar.

Por último, consideramos que los planeadores de ciudades deben tener en cuenta la importancia de la construcción de calles, ya que estas incrementan la probabilidad de construcción y la cercanía de los nuevos asentamientos al centro de la ciudad.

## 5 Conclusiones

Se analizó la predicción de la expansión urbana de la zona Zacatecas-Guadalupe usando un modelo de SVM penalizado apoyado del método SHAP. Los resultados muestran que el uso de suelo tiene un impacto alto, siendo las áreas de vegetación cercanas a las zonas urbanas y los suelos desnudos más probables de ser urbanizados, al igual que las áreas cercanas a zonas ya urbanizadas. Por otro lado, la pendiente del terreno y su dirección no tienen mucha influencia en el resultado del modelo.

Consideramos que la predicción puede ayudar a la planeación de la ciudad, principalmente evitando el descontrol del tráfico, debido al impacto que tienen las calles dentro del modelo y la forma de la ciudad.

A pesar de que el modelo refleja la interacción de las variables y cómo la ciudad se expande, consideramos que existen variables que no se exploraron, las cuales pueden aumentar la exactitud de los modelos. Como trabajo futuro, se consideraría incluir variables de proyección de población, valor de los predios y otras variables culturales que demuestren una relación más fuerte. Así como realizar un análisis comparativo con otros algoritmos como bosques aleatorios y boosting categórico.

## Agradecimientos

El autor principal agradece el apoyo recibido por el Consejo Nacional de Humanidades Ciencias y Tecnologías (CONAHCYT) a través del Programa de Becas para Estudios de Posgrado en México

## Referencias

1. Zubicaray G, Brito M, Ramírez Reyes L, et al (2021) Las ciudades mexicanas: tendencias de expansión y sus impactos. Coalition for Urban transitions: London, UK, y Washington, DC. Disponible en: <https://urbantransitions.global/publications/>.
2. Mexico. Secretaría de Desarrollo Social., Mexico. Secretaría de Gobernación., Consejo Nacional de Población (Mexico) Catálogo : sistema urbano nacional 2012, 2012th ed
3. Deng Y, Fu B, Sun C (2018) Effects of urban planning in guiding urban growth: Evidence from Shenzhen, China. *Cities* 83:118–128. <https://doi.org/10.1016/j.cities.2018.06.014>

4. Rangel CG, Soto TC, Mata V, Jiménez López E (2022) Un modelo de expansión urbana no estacionario en el espacio: Autómatas Celulares y Regresión Geográficamente Ponderada
5. Kim M, Kim G (2022) Modeling and Predicting Urban Expansion in South Korea Using Explainable Artificial Intelligence (XAI) Model. *Applied Sciences (Switzerland)* 12:. <https://doi.org/10.3390/app12189169>
6. González-Madrigal J, Solano-Lamphar H, Ramírez M (2020) La contaminación lumínica como aproximación a la planeación urbana de ciudades mexicanas
7. Nugroho F, Al-Sanjary OI (2018) A review of simulation urban growth model. *International Journal of Engineering and Technology(UAE)* 7:17–23. <https://doi.org/10.14419/ijet.v7i4.11.20681>
8. Wang SW, Munkhnasan L, Lee WK (2021) Land use and land cover change detection and prediction in Bhutan's high altitude city of Thimphu, using cellular automata and Markov chain. *Environmental Challenges* 2:. <https://doi.org/10.1016/j.envc.2020.100017>
9. Kara C, Doratlı N (2021) Predict and simulate sustainable urban growth by using GIS and MCE based CA. Case of Famagusta in Northern Cyprus. *Sustainability (Switzerland)* 13:. <https://doi.org/10.3390/su13084446>
10. Liang X, Liu X, Li X, et al (2018) Delineating multi-scenario urban growth boundaries with a CA-based FLUS model and morphological method. *Land Use Urban Plan* 177:47–63. <https://doi.org/10.1016/j.landurbplan.2018.04.016>
11. Chaturvedi V, de Vries WT (2021) Machine Learning Algorithms for Urban Land Use Planning: A Review. *Urban Science* 5
12. Li X, Gong P (2016) Urban growth models: progress and perspective. *Sci Bull (Beijing)* 61:1637–1650. <https://doi.org/10.1007/s11434-016-1111-1>
13. Karimi F, Sultana S, Shirzadi Babakan A, Suthaharan S (2019) An enhanced support vector machine model for urban expansion prediction. *Comput Environ Urban Syst* 75:61–75. <https://doi.org/10.1016/j.compenvurbsys.2019.01.001>
14. Mirbagheri B, Alimohammadi A (2018) Integration of local and global support vector machines to improve urban growth modelling. *Canadian Historical Review* 7:. <https://doi.org/10.3390/ijgi7090347>
15. Karimi F, Sultana S, Babakan AS, Suthaharan S (2021) Urban expansion modeling using an enhanced decision tree algorithm. *Geoinformatica* 25:715–731. <https://doi.org/10.1007/s10707-019-00377-8>
16. Asif M, Kazmi JH, Tariq A, et al (2023) Modelling of land use and land cover changes and prediction using CA-Markov and Random Forest. *Geocarto Int* 38:. <https://doi.org/10.1080/10106049.2023.2210532>
17. Pérez OR, Sánchez MAL, Camacho XC, Robledo MA (2020) Methodology for mining economic assimilation in Zacatecas, Mexico. *Economía, Sociedad y Territorio* 20:241–272. <https://doi.org/10.22136/est20201415>
18. Garbutt OP (2024) *Metrópolis de México 2020* Primera edición
19. Garrocho C, Soto TC, Jiménez López E *Autómata Celular Metro-: laboratorio experimental de expansión urbana*

20. Garrocho C, Jiménez E, Chávez-Soto T (2020) Expansión de la ciudad: un instrumento de simulación de escenarios para los sectores público y privado
21. Diversidad. Zacatecas. <https://cuentame.inegi.org.mx/monografias/informacion/zac/poblacion/diversidad.aspx>. Accessed 10 May 2024
22. Abrams Michael, Crippen Robert (2019) ASTER GDEM V3 (ASTER Global DEM) User Guide
23. Instituto Nacional de Estadística y Geografía XII Censo General de Población y Vivienda 2000. <https://www.inegi.org.mx/programas/ccpv/2000/>. Accessed 21 Mar 2024
24. Instituto Nacional de Estadística y Geografía Censo de Población y Vivienda 2010. <https://www.inegi.org.mx/programas/ccpv/2010/>. Accessed 21 Mar 2024
25. Instituto Nacional de Estadística y Geografía Censo de Población y Vivienda 2020. <https://www.inegi.org.mx/programas/ccpv/2020/>. Accessed 21 Mar 2024
26. González R, Barrientos A, Toapanta M, Del Cerro J (2017) Aplicación de las Máquinas de Soporte Vectorial (SVM) al diagnóstico clínico de la Enfermedad de Párkinson y el Temblor Esencial. *RIAI - Revista Iberoamericana de Automática e Informática Industrial* 14:394–405. <https://doi.org/10.1016/j.riai.2017.07.005>
27. Awad Mariette, Khanna Ragul (2015) *Efficient Learning Machines*
28. Panesar A (2020) *Machine learning and AI for healthcare: Big data for improved health outcomes*. Apress Media LLC
29. Pagano TP, Loureiro RB, Lisboa FVN, et al (2023) Context-Based Patterns in Machine Learning Bias and Fairness Metrics: A Sensitive Attributes-Based Approach. *Big Data and Cognitive Computing* 7:. <https://doi.org/10.3390/bdcc7010027>
30. Lundberg SM, Allen PG, Lee S-I (2017) *A Unified Approach to Interpreting Model Predictions*
31. Rana MS, Sarkar S (2021) Prediction of urban expansion by using land cover change detection approach. *Heliyon* 7:. <https://doi.org/10.1016/j.heliyon.2021.e08437>
32. Frimpong BF, Molkenthin F (2021) Tracking urban expansion using random forests for the classification of landsat imagery (1986–2015) and predicting urban/built-up areas for 2025: A study of the kumasi metropolis, ghana. *Land (Basel)* 10:1–21. <https://doi.org/10.3390/land10010044>