

Predicción de la generación de residuos sólidos urbanos en la Ciudad de México

Ester Calderón-Casanova¹, Mariana López-Ortíz¹, Patricia Galán¹,
Esaú Villatoro-Tello^{1,2}, Raúl R. García-Aguilar^{1,2}, Brenda García-Parra^{1,2}

¹ Universidad Autónoma Metropolitana (UAM) Unidad Cuajimalpa,
Maestría en Diseño, Información y Comunicación,
México

² Universidad Autónoma Metropolitana (UAM) Unidad Cuajimalpa,
División de Ciencias de la Comunicación y Diseño,
México

{estercalderon7,maris1589,patriciag.lara}@gmail.com
{evillatoro,rgarcia,bgarcia}@correo.cua.uam.mx

Resumen. La gestión de los residuos sólidos urbanos (RSU) en la Ciudad de México tiene fundamento en normas federales y locales, donde se señala a las delegaciones como las responsables del manejo integral de los RSU. La política del gobierno de la CDMX va encaminada hacia la prevención y minimización de los RSU a través del diseño de planes y programas para el manejo de los mismos. En este sentido, resulta necesario contar con instrumentos de apoyo que permitan conocer y entender el fenómeno, así como sus características, para el diseño de soluciones. El objetivo central de este trabajo fue desarrollar un modelo de predicción de RSU a través de técnicas de Inteligencia Artificial. Para la realización de los experimentos se buscó, identificó y extrajo información de fuentes institucionales. Los resultados obtenidos muestran que los modelos de regresión generados predicen de manera aceptable los niveles de RSU de la CDMX.

Palabras clave: Residuos sólidos urbanos, gobierno electrónico, inteligencia artificial, aprendizaje automático, modelos de regresión.

Towards the Prediction of Urban Solid Waste Generation in Mexico City

Abstract. Mexico City's (CDMX) urban solid waste (USW) management is driven through local and federal laws, which make counties responsible for their own USW management. CDMX's public policies aim to prevent and reduce the USW generation by means of designing prevention plans and sustainable programs. In this context, local government decision makers can take advantage of support tools to facilitate the implementation of strategic approaches on this regard. In this paper, our

main goal was to develop a predictive USW model using Artificial Intelligence techniques. We collected a large data set regarding the USW management plans from the CDMX. For our experiments we worked with two well-know regression techniques. Performed experiments indicated that USW prediction is possible using a small amount of features.

Keywords: Urban solid waste, e-governance, artificial intelligence, machine learning, regression models.

1 Introducción

Los residuos sólidos urbanos (RSU) son todos aquellos que se generan de actividades humanas, específicamente domésticas y comerciales en comunidades de todos los tamaños, desde aquellas con características urbanas hasta las rurales. El incremento constante en la generación de los RSU está asociado estrechamente al aumento de la población y representa un problema a nivel mundial debido a las afectaciones que ocasiona en la salud pública, la contaminación al medio ambiente en términos de tierra, aire, agua y la explotación desmesurada de recursos naturales [15].

México no se excluye de tal situación, de acuerdo con el Instituto Nacional de Geografía Estadística e Informática al 2016 se generan casi 104 mil toneladas diarias de RSU. Cabe señalar que el incremento en la población, y actividad económica son los principales factores asociados a la generación de RSU. En ese sentido la Ciudad de México (CDMX), conformada por 16 delegaciones, concentra el 30 % de la población urbana nacional, es considerada como la tercera aglomeración urbana más habitada del mundo además que representa el principal centro político, económico, científico y cultural del país; esto combinado con la gran cantidad de actividades, sitios de interés, situación laboral, etc., otorgan a la ciudad un ritmo de vida acelerado que se ve con una repercusión directa en la generación de RSU [3], representando el 13.44 % de la generación nacional [8].

Ante tales circunstancias la CDMX enfrenta un reto importante en cuanto a la Gestión Integral de Residuos Sólidos (GIRS), y en consecuencia el desarrollo de instrumentos de planeación que permitan a las delegaciones tomar mejores decisiones respecto a los requerimientos y necesidades de la población y su medio ambiente. Al respecto, organizaciones internacionales recomiendan el uso y aprovechamiento de datos en el tema para la toma de decisiones. Entre ellas, las Naciones Unidas³ (ONU) a través del Programa 21, que estipula los datos como un requisito esencial para poder seguir de cerca los cambios en cantidad y tipo de residuos y sus consecuencias para la salud y el medio ambiente. Del mismo modo, la Organización Económica para la Cooperación y el Desarrollo⁴ (OECD) recolecta, analiza y explora datos relativos a diversas cuestiones medioambientales, incluyendo los residuos sólidos, con el fin de predecir tendencias a futuro

³ www.un.org/spanish/esa/sustdev/agenda21/agenda21spchapter21.htm

⁴ <https://www.oecd.org/about/>

y proporcionar a los gobernantes entendimiento sobre las problemáticas y sus cambios en busca de soluciones.

A pesar de los esfuerzos que se realizan en distintos países, y las observaciones y recomendaciones hechas por organizaciones de carácter mundial, la CDMX aún carece de herramientas que permitan el aprovechamiento y explotación de la información disponible para gestionar los RSU. Así entonces, en este trabajo se propone y evalúa un modelo de Inteligencia Artificial (IA) que realiza la predicción de la generación de RSU dentro de la CDMX. Nuestro modelo se entrenó a partir de información que fue recolectada de fuentes institucionales, específicamente la SEDEMA⁵. Los experimentos realizados muestran que es posible entrenar modelos automáticos para la predicción de RSU a partir de un conjunto acotado de variables (atributos), mismas que son ya actualmente recolectadas y documentadas por las autoridades de la CDMX. Además de esto, el análisis realizado en este proceso de minería de datos mostró que la generación de RSU depende de aspectos que van más allá del aumento en la población, indicando una fuerte correlación con variables que no eran evidentes hasta el momento.

El resto del artículo se organiza de la siguiente manera. La sección 2 describe el trabajo relacionado, en la sección 3 se describe el proceso de recolección de datos con los cuales se entrenó y evaluó nuestro modelo de predicción. La sección 4 describe la metodología empleada para la realización de los experimentos, mientras que la sección 5 describe los experimentos y resultados obtenidos. Finalmente, la sección 6 plantea las conclusiones obtenidas y las líneas de trabajo futuro.

2 Trabajo relacionado

Actualmente, existe un creciente interés tanto en países de primer mundo como en países en vías de desarrollo por aprovechar las técnicas de inteligencia artificial en predicción de la generación de RSU [7,5,6,12,4,1,9,11,13,2]. A continuación se da una breve descripción de los trabajos más recientes y que se asemejan en gran medida a la propuesta aquí planteada. Es importante mencionar que debido a que no existe una base de datos estándar con la cual sistemas de predicción de RSU puedan ser evaluados, resulta difícil tener una comparación directa entre los distintos esfuerzos realizados hasta el momento.

En [6] los autores desarrollan un modelo de optimización de regresión basado en máquinas de vectores de soporte para la planificación de la gestión de residuos sólidos municipales (RSM) en los distritos urbanos de Beijing, China. El modelo desarrollado no solo puede predecir la cantidad futura de generación de residuos de la ciudad, sino que también refleja características dinámicas, interactivas e inciertas del sistema de gestión de RSU. Los autores evaluaron cuatro funciones de kernel; lineal, polinomial, base radial y un perceptrón de capa múltiple. En sus experimentos, el kernel polinomial les permite obtener los mejores resultados.

⁵ <http://www.sedema.cdmx.gob.mx/>

Por otro lado, en [4] los autores presentan un análisis sobre las variables que inciden en la estimación de residuos municipales recolectados en Venezuela, para lo cual emplearon técnicas de análisis multivariado, por ejemplo, correlación de Pearson. Para su análisis, emplearon información de 175 municipios, identificando que atributos como son el tamaño de la población urbana, número y tipo de vivienda son variables altamente correlacionadas con la generación de RSU. De forma similar, en [9] los autores evaluaron la relevancia de variables como son: número de residentes, edad de la población, esperanza de vida urbana, en un modelo de predicción de residuos sólidos municipales en Iasi Rumania. Al contrario de otros trabajos, en [9] evalúan la predicción de RSU a un nivel más fino, es decir, buscan generar modelos de predicción por tipo de RSU, por ejemplo, papel, plástico, metal, vidrio, biodegradables y otros residuos.

En [1] los autores evaluaron cuatro modelos de inteligencia artificial como máquinas de vectores soporte (SVM), sistema adaptativo de inferencia neurodifusa (ANFIS), redes neuronales (ANN), y vecinos más cercanos (kNN). El objetivo fue determinar el mejor modelo para predecir mensualmente la generación de residuos en el Ayuntamiento de Logan en Queensland Australia. De sus experimentos, los autores reportan que el modelo denominado ANFIS fue el que mejor comportamiento obtuvo alcanzando un $R^2 = 0.98$.

Más recientemente, en [11] los autores emplean información de los censos municipales de Ontario, Canada. Las variables que los autores consideran son de información demográfica y socio-económica que describe a los municipios considerados (220 en total). Para la evaluación de sus modelos emplearon técnicas basadas en árboles de decisión y redes neuronales (ANN), llegando a la conclusión de que las ANN son mejores predictores alcanzando un $R^2 = 0.86$. Por otro lado, en [13] los autores evalúan la importancia de variables como son el tamaño de la población, la edad, los ingresos y el número de turistas en un modelo de predicción de RSU. Los resultados que alcanzan reportan valores de $R^2 = 0.96$ empleando técnicas de redes neuronales.

Finalmente, en un estudio más ambicioso reportado en [2], los autores evaluaron el comportamiento de un modelo general de regresión basado en redes neuronales para la predicción de RSU en 44 países. Contrario a los trabajos descritos anteriormente, este artículo tuvo como principal objetivo determinar el impacto de las crisis económicas en la generación de RSU. Para sus experimentos emplearon tanto indicadores socio-económicos, datos demográficos, así como indicadores de planes de sustentabilidad.

Como es posible observar, muchos trabajos reportan resultados alentadores en sus modelos de predicción implementados. Sin embargo, nótese que no hay una evidencia clara sobre qué método de regresión puede ser el mejor, pues los resultados de la predicción dependen en gran medida de los datos disponibles. Varios de estos trabajos previos coinciden en la dificultad que representa tener datos confiables para la realización de estos ejercicios, por lo mismo no hay una evidencia clara de qué factores (variables o indicadores) son los más relevantes para la construcción de los modelos de predicción. Inspirados por estos esfuerzos, nuestro trabajo propone y evalúa un modelo de inteligencia artificial que realiza

la predicción de la generación de RSU dentro de la CDMX, el cual, contrario a los trabajos previos, emplea información sobre gestión de los RSU que es recolectada por las distintas delegaciones de la ciudad, con el objetivo de determinar la pertinencia de estos indicadores en el proceso de predicción. Agregado a lo anterior, es conveniente mencionar que hasta donde sabemos, este tipo de trabajos no se ha realizado previamente en la CDMX, por lo cual, este artículo representa un trabajo innovador en muchos sentidos para el contexto nacional.

3 Datos

Uno de los esfuerzos más relevantes para la realización de este trabajo fue la búsqueda y recolección de datos relativos a RSU en México, específicamente en la CDMX. Los datos, como insumo y núcleo del proyecto, resultan muy valiosos debido a la escasez de éstos, así como su función para representar y visibilizar el problema de la gestión de los RSU. Varios trabajos [4,2] reportan la misma situación, la falta de datos en el tema, sin embargo esto no ha sido obstáculo para la formulación de propuestas de solución utilizando enfoques de inteligencia artificial. Motivados por estos esfuerzos previos, este artículo tiene como finalidad determinar hasta que punto es posible entrenar un modelo de predicción de RSU en el contexto nacional, específicamente para la CDMX.

El primer paso para la obtención de datos fue la ubicación de las fuentes institucionales que los generan y publican. Se identificaron diversas organizaciones a nivel nacional como INEGI, SEGOB y CONAPO⁶ las cuales cuentan con información relacionada, por ejemplo, generación de RSU, población e ingreso per capita a nivel municipal y delegacional. Sin embargo, para la realización de nuestro estudio se utilizaron los datos proporcionados por la Secretaría del Medio Ambiente (SEDEMA) de la CDMX.

Dichos datos, proporcionado por este organismo, contiene características que la hacen muy valiosa y útil para nuestros propósitos, por ejemplo su periodicidad, diversidad y cantidad de información relativa a RSU. Es importante mencionar que la SEDEMA inició desde el 2006 el monitoreo de información relacionada a la gestión de los RSU en las 16 delegaciones de la CDMX. Con esta información se genera una publicación anual, denominada Inventarios de Residuos Sólidos de la Ciudad de México. Para el trabajo presentado en este artículo se utilizó la información extraída de estos inventarios.

En la tabla 1⁷ se muestra el inventario de variables monitoreadas por la SEDEMA para cada una de las 16 delegaciones de la CDMX, así como los años para los cuales existe información de dichas variables ('-' no existe información, 'x' si existe información). En total son 39 variables que refieren a distintos aspectos de la gestión de los RSU como son: equipo, infraestructura, personal, planes de manejo y generación de RSU, puntos de recolección, etc.

⁶ <https://www.gob.mx/conapo>

⁷ Es importante mencionar que existen muchas más variables, sin embargo para los experimentos reportados en este trabajo sólo se emplearon las variables que tienen información desglosada a nivel delegacional.

A pesar de la validez de esta información, observe que no todas las variables tienen información para los años considerados, i.e., 2006-2016. La razón a la que atribuimos esta ausencia de información es debido a que las distintas gestiones de la CDMX han decidido prestar atención a aspectos diferentes, por lo cual algunas variables desaparecen mientras que otras van surgiendo. El resultado de esto es la carencia de información en gran parte de las variables entre los años 2006 y 2010 principalmente, sin embargo note que la variable de interés, i.e., la que interesa modelar (“Generación de residuos sólidos (ton/día)”) está presente para todos los años considerados.

En su totalidad, la base de datos que se logró compilar contiene 176 instancias (16 delegaciones \times 11 años) representadas a través de 39 atributos (variables), es decir un total de 6864 registros.

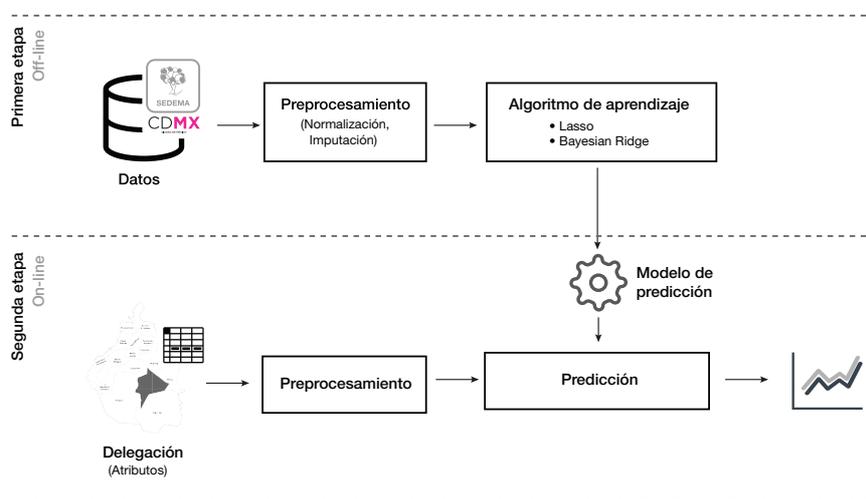


Fig. 1. Metodología propuesta para la predicción de la generación de RSU.

4 Metodología propuesta

La figura 1 muestra de manera esquemática la arquitectura del método propuesto para hacer la predicción del valor de generación de RSU. Contrario a un problema de clasificación, nuestro objetivo implica determinar un valor continuo, el valor de “Generación de residuos sólidos (ton/día)”; por lo tanto abordamos el reto como un problema de regresión. A continuación describimos en forma breve cada uno de los módulos de nuestra metodología empleada.

Tabla 1. Inventario y clasificación por año de variables identificadas. *PES (Puntos específicos de recolección separada). *PMGRS (Planes de manejo y generación de residuos sólidos).

Variable	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Rutas con recolección separada	x	x	x	x	x	x	x	x	-	-	-
Rutas totales de recolección	x	x	x	x	x	x	x	x	x	x	x
Colonias con separación	x	x	x	x	x	x	x	x	-	-	-
Colonias totales por delegación	x	x	x	x	x	x	x	-	x	x	x
Planes de manejo	x	x	x	-	x	x	x	x	x	x	x
Generación de residuos sólidos (ton/día)	x	x	x	x	x	x	x	x	x	x	x
Eficiencia de la separación de r. orgánicos	-	-	-	-	-	-	x	x	-	-	-
Eficiencia de la recolección de r. orgánicos	-	-	-	-	-	-	-	-	x	x	x
Total de vehículos recolectores	-	x	x	x	x	x	x	x	x	x	x
Tipo de vehículo: Carga trasera	-	-	-	x	-	-	x	x	x	x	x
Tipo de vehículo: Rectangular	-	-	-	-	-	-	x	x	x	x	x
Tipo de vehículo: tubular	-	-	-	-	-	-	x	x	x	x	x
Tipo de vehículo: Volteo	-	-	-	-	-	-	x	x	x	x	x
Tipo de vehículo: Frontal	-	-	-	-	-	-	-	x	x	x	x
Tipo de vehículo: Doble compartimento	x	x	x	x	x	x	x	x	x	x	x
Otros vehículos	-	-	-	x	-	x	x	x	x	x	x
Barrido manual número de barredores	-	-	-	-	-	-	-	x	x	x	x
Barrido manual número de carritos	-	-	-	-	-	-	-	x	x	x	x
Barrido manual número de rutas o tramos	-	-	-	-	-	-	-	x	x	x	x
Número de barredoras mecánicas	-	-	-	-	-	-	-	x	x	x	x
Chóferes	-	-	-	-	-	-	-	x	x	x	x
Voluntarios	-	-	-	-	-	-	-	x	x	x	x
PES en unidades habitacionales	-	-	-	-	-	-	x	x	x	x	x
PES en escuelas	-	-	-	-	-	-	x	x	x	x	x
PES en mercados	-	-	-	-	-	-	x	x	x	x	x
PES centro comercial	-	-	-	-	-	-	x	x	x	x	x
PES en terminales	-	-	-	-	-	-	x	x	x	x	x
PES en parques y plazas	-	-	-	-	-	-	x	x	x	x	x
PES de servicios e industria	-	-	-	-	-	-	x	x	x	x	x
PES Otro	-	-	-	-	-	-	x	x	x	x	x
PMGRS por categoría A	-	-	-	-	-	x	x	x	x	x	x
PMGRS por categoría B	-	-	-	-	-	x	x	x	x	x	x
PMGRS por categoría C	-	-	-	-	-	x	x	x	x	x	x
PMGRS por categoría D	-	-	-	-	-	x	x	x	x	x	x
PMGRS por categoría E	-	-	-	-	-	x	x	x	x	x	x
PMGRS por sector comercio	-	-	-	-	-	-	x	x	x	x	x
PMGRS por sector industria	-	-	-	-	-	-	x	x	x	x	x
PMGRS por sector servicios	-	-	-	-	-	-	x	x	x	x	x

4.1 Preprocesamiento

Como se describió en la sección 3, la base de datos empleada está conformada con información proveniente de la SEDEMA. La tabla 1 muestra la existencia

cia/ausencia (x/-) de los 39 indicadores (atributos) que han sido monitoreados en el periodo de 2006-2016. Nótese un problema importante de datos faltantes, específicamente de los 6864 registros, 3323 son vacíos, es decir, el 48 %.

Tener una cantidad importante de datos faltantes representa un reto para todo algoritmo de aprendizaje automático. Por tal razón, como parte del pre-procesamiento de los datos se aplicaron técnicas de imputación. El proceso de imputación refiere a remplazar datos faltantes por un valor sustituto. Para los experimentos reportados en este trabajo se utilizó como técnica de imputación de cada atributo k el valor de la media (\bar{x}_k).

Como operación adicional del pre-procesamiento de los datos se aplicó un proceso de normalización. Específicamente se aplicó la norma l^2 o norma euclidiana, la cual se calcula como se observa en la ecuación (1):

$$|\mathbf{X}| = \sqrt{\sum_{k=1}^n |x_k|^2}, \quad (1)$$

donde x_k representa al atributo k y n es el número total de atributos. Así entonces, al final el valor de cada atributo k se re-calcula como $x_k = \frac{x_k - \bar{x}_k}{|\mathbf{X}|}$.

4.2 Métodos de regresión

Los métodos de regresión tienen como finalidad principal construir un modelo que sea capaz de predecir el valor cuantitativo de una variable a partir de un conjunto de datos históricos. Al final, el modelo será capaz de hacer una predicción sobre un nuevo dato del cual se desconoce su respuesta. En otras palabras, el modelo predictivo puede ser descrito como el problema de aproximar una función f con variables de entrada X a su correspondiente valor continuo de salida y . A esto se le conoce como el problema de encontrar la función de aproximación $f(X) = y$.

A la variable y se le conoce como la variable (cuantitativa) de respuesta, mientras que X es el conjunto de atributos o variables predictores. Así entonces, lo que se busca es encontrar los coeficientes β que satisfacen a: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$. Para los experimentos realizados en este trabajo empleamos como métodos de regresión el algoritmo de Lasso [14] y Bayesian-Ridge [10].

Lasso. Es un tipo de regresión lineal que utiliza técnicas de “encogimiento” (shrinkage), el cual tiene como objetivo obtener el conjunto de predictores que minimizan el error de predicción de la variable de respuesta (ecuación (2)). El encogimiento significa que los valores de los predictores son llevados a un punto central, por ejemplo la media o incluso cero:

$$\sum_{i=1}^n (y_i - \sum_j x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|. \quad (2)$$

Contrario a un método de mínimos cuadrados ordinarios (OLS), Lasso impone restricciones en el encogimiento, específicamente penalizando por medio

de considerar el valor absoluto de la magnitud de los coeficientes, en otras palabras por el valor de la norma l^1 (segundo término de la ecuación (2)).

Bayesian Ridge. El método de Ridge es muy similar al método de regresión de Lasso, la diferencia radica principalmente en la penalización por medio de considerar el cuadrado de la magnitud de los coeficientes (vea ecuación (3)), en otras palabras, la norma l^2 .

$$\sum_{i=1}^n (y_i - \sum_j x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j^2|. \quad (3)$$

El aspecto probabilístico (bayesian) de este método es debido a que los valores de los coeficientes son estimados a priori. Para nuestros experimentos, los valores a priori son estimados bajo una distribución Gaussiana. Este tipo de técnica es recomendada cuando existen una insuficiencia de datos.

En general, ambos métodos de regresión se recomiendan cuando los datos muestran altos niveles de correlación, el cual es el nuestro caso. Para los experimentos realizados se utilizó la implementación de Lasso y Bayesian Ridge disponibles en scikitlearn⁸.

4.3 Evaluación

Para la evaluación del desempeño del modelo de predicción se utilizaron métricas estándar, como son el error medio absoluto (MAE), el coeficiente de determinación (R^2), y la varianza (EV), las cuales se calculan como se muestra en las ecuaciones (4), (5) y (6) respectivamente:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (5)$$

$$\text{EV} = 1 - \frac{\text{Var}\{y - \hat{y}\}}{\text{Var}\{y\}}, \quad (6)$$

donde n es el número de instancias, \hat{y} el el valor predicho por el modelo, y es el valor real, \bar{y} es el valor medio de y , y Var es la varianza estadística σ^2 . Tanto EV como R^2 son métricas que, entre más cercano a 1.0, significa que mejor es el desempeño del modelo de predicción. Por otro lado, MAE es una medida que crece conforme más errores comete el método de predicción, por lo cual, un valor cercano a 0 es preferible.

Finalmente, es conveniente mencionar que para la realización de los experimentos se utilizó como técnica de validación una estrategia de validación cruzada de 10 pliegues.

⁸ <http://scikit-learn.org/>

Tabla 2. Resultados obtenidos de los experimentos realizados.

Experimento	Algoritmo	#Inst	#Atts	Prep-Op.	MAE	R^2	V
EXP-1	Lasso	176	39	l^2	281.26	0.40	0.40
	B-Ridge			l^2	267.62	0.45	0.45
	Lasso			l^2, \bar{x}_k	258.53	0.52	0.52
	B-Ridge			l^2, \bar{x}_k	266.97	0.52	0.52
EXP-2	Lasso	64	39	l^2	152.45	0.78	0.78
	B-Ridge			l^2	181.50	0.68	0.68
	Lasso			l^2, \bar{x}_k	181.50	0.68	0.68
	B-Ridge			l^2, \bar{x}_k	204.37	0.64	0.64
EXP-3	Lasso	64	35	l^2	159.89	0.76	0.76
	B-Ridge			l^2	208.65	0.65	0.65
	Lasso			l^2, \bar{x}_k	187.37	0.64	0.64
	B-Ridge			l^2, \bar{x}_k	220.44	0.61	0.61
EXP-4	Lasso	64	10	l^2	160.71	0.79	0.79
	B-Ridge			l^2	166.67	0.77	0.77

5 Experimentos y resultados

En esta sección se describen los experimentos realizados así como los resultados obtenidos. Cuatro grandes experimentos fueron realizados:

- **EXP-1.** El objetivo de este experimento fue comprobar el impacto de entrenar un modelo de predicción empleando toda la base de datos disponible.
- **EXP-2.** El objetivo de este experimento fue evaluar el impacto de eliminar de la base de datos los años que presentaban el porcentaje más alto de datos faltantes, i.e., 2006 a 2012.
- **EXP-3.** En este experimento se evaluó el impacto de eliminar aquellas variables ausentes del conjunto de datos resultante del EXP-2; en total 3 variables son descartadas (Total de rutas con recolección separada, total de colonias con recolección separada y eficiencia de la separación de los residuos orgánicos por delegación).
- **EXP-4.** Para este experimento se midió la correlación de Pearson entre las 35 variable restantes contra la variable de interés (i.e., Generación de residuos sólidos (ton/día)). Para la construcción de la representación conservamos sólo aquellas variables con correlación mayor a 0.5,

Los resultados de los experimentos se muestran en la tabla 2. La segunda columna especifica el algoritmo de regresión empleado, la columna **#Inst** indica el número de instancias presentes en el experimento, **#Atts** refiere el número de atributos con que cada instancia es representada, **Pre-Op** indica las operaciones aplicadas a los datos, como son normalización (l^2) e imputación (\bar{x}_k).

En primer lugar, observe que tener un porcentaje tan alto de datos vacíos (EXP-1) impacta de forma negativa en el desempeño del método de regresión. Nótese que cuando no se realiza imputación, el desempeño obtenido es de un $R^2 = 0.45$ con el método de B-Ridge. Por otro lado, cuando se hacen operaciones

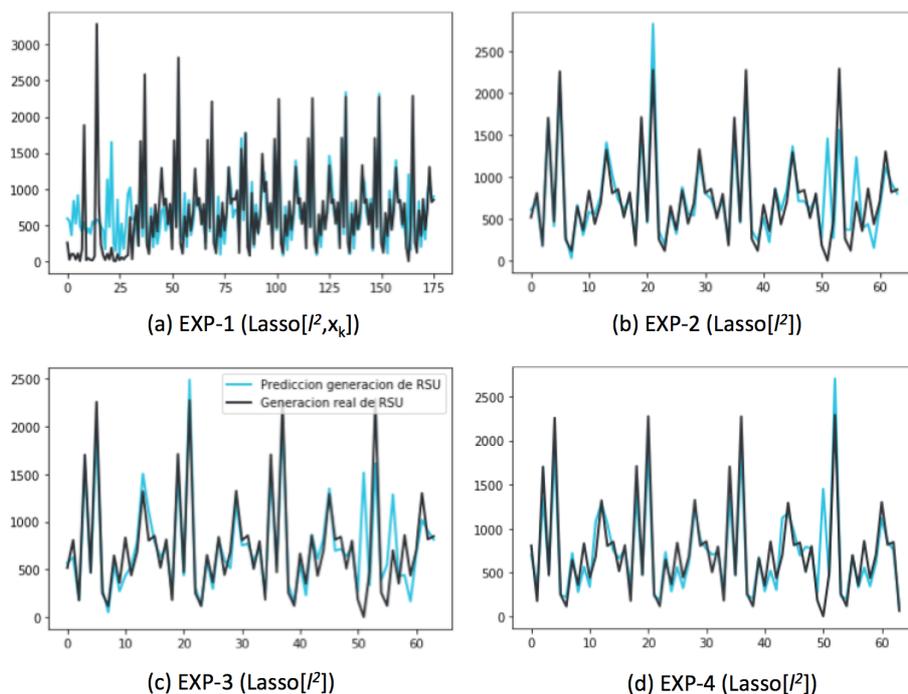


Fig. 2. Gráficas de comparación entre las mejores configuraciones de cada experimento realizado. El eje de las y 's representa el valor de RSU, mientras que el eje de las x 's representa las instancias, las cuales están ordenadas cronológicamente.

de imputación el coeficiente de determinación sube a $R^2 = 0.52$ con el método Lasso. En segundo lugar, los resultados de los experimentos EXP-2 y EXP-3 muestran que haber hecho la eliminación de las instancias de los años 2006 a 2012 permite al modelo de regresión obtener un mejor desempeño ($R^2 = 0.78$). La eliminación de estas instancias se justifica debido a que la gran mayoría de ellas tenían ausencia de datos para gran cantidad de variables. Finalmente, los resultados obtenidos del EXP-4 indican que hacer una selección de atributos permite mejorar el valor del coeficiente de determinación ($R^2 = 0.79$). Este resultado indica, hasta cierto punto, que es suficiente emplear solo diez atributos para lograr tener un modelo de predicción de RSU con un desempeño aceptable.

En la figura 2 se muestra gráficamente el desempeño de cada uno de los experimentos realizados en su mejor configuración. La línea más oscura (negra) denota el valor real de RSU mientras que la línea más tenue (azul) es el valor que se predice. Observe para el EXP-1 (a), las primeras instancias, las cuales corresponderían a los años de 2006 a 2009 aproximadamente, son las que peor comportamiento tienen. Por otro lado, el experimento que mejor desempeño tuvo es el (d) EXP-4, donde se puede observar que el comportamiento del modelo de predicción es mejor que (b) y (c), sobre todo hacia las últimas instancias.

6 Conclusiones

En este artículo se describe un método para la predicción de residuos sólidos urbanos dentro de la CDMX. El método propuesto busca impactar de manera positiva en un problema de relevancia mundial, i.e., la generación de RSU. Hasta donde se sabe, nuestro trabajo es pionero en buscar aplicar técnicas de IA para atender este problema dentro del contexto nacional y en particular en una de las urbes más pobladas del mundo, la Ciudad de México.

Los experimentos realizados muestran que es posible hacer una predicción adecuada de los valores de RSU, alcanzando desempeños de $R^2 = 0.79$. Como parte de los experimentos, se identificó un conjunto de 10 variables con las cuales se puede hacer esta predicción. Contrario al trabajo reportado en la literatura, nuestro método no emplea información socio-económica ni demográfica para hacer la representación de los municipios/delegaciones, con lo cual se muestra que hay otros factores involucrados en la generación de RSU. El modelo desarrollado puede servir para unidades territoriales similares a las delegaciones de la Ciudad de México, por ejemplo como son todos los municipios del país, siempre y cuando cuenten con la información mínima de las 10 variables/atributos requeridas para la predicción.

Es conveniente reflexionar sobre la importancia que tiene la disponibilidad de información. Es necesario resaltar la obligación de los gobiernos actuales para invertir en iniciativas que les permita monitorear de forma periódica aspectos relevantes a la gestión de RSU. Contar con mecanismos constantes permitirá generar información que eventualmente podría favorecer a los métodos de predicción. En este mismo sentido, el análisis de los datos obtenidos a través de la SEDEMA permitió identificar los cambios de intereses entre distintos periodos de gestión de la CDMX. Esto se vio reflejado debido a la aparición/desaparición de distintas variables a lo largo de los años. Esto evidencia, hasta cierto punto, una falta de planificación a largo plazo.

Como trabajo futuro haremos más experimentos incorporando información del tipo socio-demográfico, económico y/o de educación. En el trabajo previo reportado hasta el momento, se ha mostrado que este tipo de variables impacta positivamente a los modelos de regresión. Además de esto, es de nuestro particular interés integrar el modelo desarrollado en un sistema de información para la gestión de los RSU en la CDMX, el cual servirá como herramienta de apoyo en el análisis del estado actual de las delegaciones, se visualizará información sobre infraestructura, servicios y la generación de los RSU en la CDMX⁹.

Agradecimientos. El trabajo de las primeras tres autoras fue parcialmente financiado por el CONACyT a través de las becas de maestría 616127, 616489, 775648 respectivamente. Agradecemos además el apoyo otorgado por la Coordinación de la Maestría en Diseño, Información y Comunicación (MADIC) de la UAM Cuajimalpa, así como el apoyo de la División de Ciencias de la Comunicación y Diseño de la UAM Cuajimalpa.

⁹ El sistema de información estará disponible en: www.rsucdmx.com

Referencias

1. Abbasi, M., El Hanandeh, A.: Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Management* 56, 13–22 (2016)
2. Adamović, V.M., Antanasijević, D.Z., Ristić, M.Đ., Perić-Grujić, A.A., Pocaajt, V.V.: Prediction of municipal solid waste generation using artificial neural network approach enhanced by structural break analysis. *Environmental Science and Pollution Research* 24(1), 299–311 (2017)
3. del Medio Ambiente de la CDMX-SEDEMA, S.: Programa de gestión integral de los residuos sólidos 2016-2020. Tech. rep., SEDEMA (2016)
4. Chassaingne, G., Pinto, G.: Determinación de variables que inciden en la estimación de residuos y desechos sólidos municipales recolectados en venezuela. *Interciencia* 39(12) (2014)
5. Chung, S.S.: Projecting municipal solid waste: The case of hong kong sar. *Resources, Conservation and Recycling* 54(11), 759–768 (2010)
6. Dai, C., Li, Y., Huang, G.: A two-stage support-vector-regression optimization model for municipal solid waste management—a case study of beijing, china. *Journal of environmental management* 92(12), 3023–3037 (2011)
7. Dyson, B., Chang, N.B.: Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling. *Waste management* 25(7), 669–679 (2005)
8. Geografía (INEGI), I.N.d.E.y.: Censo nacional de gobiernos municipales y delegacionales 2017. CNGDM, <http://www.beta.inegi.org.mx/proyectos/censosgobierno/municipal/cngmd/2017/>
9. Ghinea, C., Drăgoi, E.N., Comăniță, E.D., Gavrilesco, M., Câmpean, T., Curteanu, S., Gavrilesco, M.: Forecasting municipal solid waste generation using prognostic tools and regression analysis. *Journal of environmental management* 182, 80–93 (2016)
10. Haitovsky, Y., Wax, Y.: Generalized ridge regression, least squares with stochastic prior information, and bayesian estimators. *Appl. Math. Comput.* 7(2), 125–154 (Sep 1980), [http://dx.doi.org/10.1016/0096-3003\(80\)90002-8](http://dx.doi.org/10.1016/0096-3003(80)90002-8)
11. Kannangara, M., Dua, R., Ahmadi, L., Bensebaa, F.: Modeling and prediction of regional municipal solid waste generation and diversion in canada using machine learning approaches. *Waste Management* (2017)
12. Srivastava, A.K., Nema, A.K.: Fuzzy parametric programming model for multi-objective integrated solid waste management under uncertainty. *Expert Systems with Applications* 39(5), 4657–4678 (2012)
13. Sun, N., Chungpaibulpatana, S.: Development of an appropriate model for forecasting municipal solid waste generation in bangkok. *Energy Procedia* 138, 907–912 (2017)
14. Tibshirani, R.: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 267–288 (1996)
15. Un-Habitat: Solid waste management in the world's cities: water and sanitation in the world's cities 2010. UN-HABITAT (2010)