

PREDICCIÓN DE LA GENERACIÓN DE RESIDUOS SÓLIDOS DOMICILIARIOS CON MACHINE LEARNING EN UNA ZONA RURAL DE PUNO

PREDICTION OF SOLID HOUSEHOLD WASTE GENERATION WITH MACHINE LEARNING IN A RURAL AREA OF PUNO

Alberto Franco Cerna-Cueva^{1*} , Cesar Rosas-Echevarría² , Roberto Perales-Flores³ 
, Pierina Lisbeth Ataucusi-Flores⁴ 

^{1,4}Universidad Nacional Agraria de la Selva, Facultad de Recursos Naturales Renovables, Escuela Profesional de Ingeniería Ambiental, Tingo María, Perú

^{2,3}Universidad Nacional Hermilio Valdizán, Facultad de Ingeniería Industrial y de Sistemas, Huánuco, Perú

Recibido (Received): 09 / 11 / 2021 Aceptado (Accepted): 25 / 06 / 2022

RESUMEN

La gestión de residuos sólidos es uno de los principales desafíos ambientales en todas las ciudades del mundo debido a factores como el crecimiento poblacional y los hábitos de consumo. Una de las principales herramientas para el diseño de proyectos de gestión de residuos, es la estimación de la generación per cápita, sin embargo, el método tradicional para obtener esta información demanda mucho esfuerzo y tiempo, por ello esta investigación plantea un enfoque alternativo de la estimación de la generación per cápita a partir de factores socioeconómicos. Para ello se recogió información socio económica demográfica e información sobre la generación per cápita de residuos sólidos de 50 familias del centro poblado "El Juncal", departamento de Puno, posteriormente se determinaron las variables que tienen influencia significativa a partir del coeficiente de correlación ρ de Spearman para las variables numéricas y un ANOVA para las variables categóricas con un umbral de aceptación de 0.4 y 0.05 respectivamente. Las variables seleccionadas se utilizaron para entrenar los modelos de redes neuronales, regresión lineal múltiple, Support Vector Machine, procesos gaussianos y Random Forest, cuyos desempeños fueron de $R^2 = 0.986, 0.982, 0.959, 0.942, 0.832$; respectivamente. Para la validación se utilizó validación cruzada y partición de datos. Los resultados indican que las variables influyentes son el ingreso per cápita, el gasto en insumos y productos, el tamaño familiar y los servicios del hogar. Se concluye que las predicciones de los modelos son confiables con valores de error cuadrático medio (RMSE) de 8 a 27 g

Palabras clave: Desperdicios, factor social, algoritmos de machine learning, gestión, suburbios, domicilio

ABSTRACT

Solid waste management is one of the main environmental challenges in all cities of the world due to factors such as population growth and consumption habits. One of the main tools for the design of waste management projects is the estimation of per capita generation, however, the traditional method to obtain this information demands a lot of effort and time, so this research proposes an alternative approach to estimate per capita generation based on socioeconomic factors. For this purpose, socioeconomic demographic information and information on the per capita generation of solid waste was collected from 50 families in the "El Juncal" population center, department of Puno, then the variables that have significant influence were determined from the Spearman's ρ correlation coefficient for numerical variables and an ANOVA for categorical variables with an acceptance threshold of 0.4 and 0.05 respectively. The selected variables were used to train neural network models, multiple linear regression, Support Vector Machine, Gaussian processes and Random Forest, whose performances were $R^2 = 0.986, 0.982, 0.959, 0.942, 0.832$; respectively. Cross validation and data partitioning were used for validation. The results indicate that the influential variables are per capita income, expenditure on inputs and outputs, family size and household services. It is concluded that the predictions of the models are reliable with root mean square error (RMSE) values of 8 to 27 g

Keywords: Waste, social factor, machine learning algorithms, management, suburbs, domicile

1. INTRODUCCIÓN

En los últimos años, la gestión de los residuos sólidos (RS) se ha convertido en uno de los mayores retos en muchas partes del mundo. Al igual que otros países en vías de desarrollo, Perú se enfrenta a un gran reto en la gestión de la creciente cantidad de RS debido a su crecimiento demográfico, la migración de la población

* Corresponding author.: Alberto F. Cerna Cueva
E-mail: Francoc_92_o8@hotmail.com

de las zonas rurales a las urbanas y muchos otros factores. La tendencia al aumento de la generación de RS es un motivo de preocupación para las zonas urbanas y rurales del Perú.

En el Perú se recolecta por día 22860,65 toneladas de RS y una persona genera en promedio 0,58 Kg de residuos (INEI 2020). La cantidad de RS generados ha aumentado considerablemente en los últimos años, en el 2014 la generación de residuos fue de 6,90 millones de toneladas a nivel nacional y en el 2019 fueron 7,78 millones de toneladas, el 90% de los residuos termina en uno de los 1850 botaderos informales [3]. Según la Dirección General de Gestión de Residuos Sólidos, al año el Perú emite 4,482 millones de toneladas de CO₂ por la disposición final de RS municipales [2]. Actualmente, la mayoría de las municipalidades del país realizan labores de manejo de RS sin valerse de información fundamental acerca de la caracterización y generación y como resultado de esto la población recibe una mala calidad del servicio, otra consecuencia de la ausencia de estos estudios es la improvisación en cuanto a las soluciones brindadas, haciendo ineficiente la gestión de residuos. Los estudios de generación de RS brindan información para el diseño técnico del recojo y la disposición final. Aunque en el Perú existen una norma para la caracterización de RS municipales (Resolución Ministerial N° 457-2018-MINAM) y el método tradicional habitual para cuantificar los residuos generados son el muestreo directo; este método es costoso y requiere mucho tiempo [4] y por razones como ésta a veces se dejan de lado los centros poblados dispersos e incluso algunos sectores en las mismas ciudades. El manejo de los residuos sólidos municipales en el Distrito de San Antonio de Esquilache en Puno, se limitan a labores parciales de recolección y transporte de los residuos sólidos del ámbito urbano, de forma improvisada, esta situación afecta directamente a los habitantes del Centro Poblado El Juncal, debido a que es en este centro poblado en donde sucede la disposición final insalubre de los RS en botaderos a cielo abierto para ser quemados [5].

2. ANTECEDENTES

Los modelos de machine learning o aprendizaje automático están siendo utilizados para la estimación y/o predicción de la generación de RS tanto municipales, industriales, hospitalarios, a partir de las técnicas de aprendizaje profundo también se puede predecir la composición de los residuos y las tendencias de aumento de cada tipo por lo que la aplicación de estas técnicas es creciente. Los modelos más utilizados para la predicción de la generación de RS a partir de factores socioeconómicos son modelos basados en redes neuronales [4], [6]–[16], las regresiones [7], [10]–[12], [15], [17], los árboles de decisión o Random Forest (RF) [6], [9], [11]–[13], [18], los procesos gaussianos [7], [12]

y los modelos de support vector machine (SVM) [6], [9], [12], por otra parte también se utilizan algoritmos de aprendizaje profundo no supervisado como k-means, para la clasificación en grupos de los tipos de generadores de RS en función a las características socioeconómicas y demográficas [19].

Con respecto a la generación de residuos domiciliarios influyen factores socioeconómicos como el ingreso mensual, el gasto, y factores demográficos como el tamaño de los hogares, el nivel de alfabetización y la cantidad de personas [20]. Los factores socioeconómicos como el tamaño de la población, la esperanza de vida, el nivel de alfabetización, el desarrollo humano y el ingreso per cápita influyeron en la cantidad de residuos generados en 39 municipios de Sao Paulo (Brasil) determinados mediante el coeficiente de correlación de Pearson [21]. El nivel educativo, el tamaño de la familia, los ingresos y el tipo de empleo influyen en la generación de residuos domésticos generados per cápita por día en la ciudad de Bangalore (India) [22]. Los datos que intervienen en la modelización de la generación de residuos están relacionados principalmente con la producción, el consumo o la eliminación de residuos [23], [24], como la población [25]–[27], los ingresos [25], la educación [28], [29], la edad [30] y el empleo [28], [30]. Aunque las experiencias expuestas anteriormente son casos exitosos de la aplicación de modelos de Machine learning para la predicción de la generación per cápita de RS, cabe mencionar que en estos estudios se resalta la necesidad de datos confiables, ya que las personas suelen ser celosas de la información de sus viviendas por lo que pueden brindar información sesgada, siendo necesario que la información obtenida para la construcción de estos modelos sean fuentes estrictamente fuentes primarias.

Por lo expuesto anteriormente esta investigación tiene los siguientes objetivos: 1) Identificar las variables socioeconómicas, demográficas numéricas y categóricas que están correlacionadas e influyen significativamente en la generación per cápita de RS en el centro poblado el Juncal y; 2) a partir de las variables seleccionadas construir y validar modelos de Machine Learning para la predicción de generación de RS per cápita

3. METODOLOGÍA

3.1 Ubicación

La investigación se llevó a cabo en el centro poblado del Juncal que se encuentran en el distrito de San Antonio, provincia de Puno, departamento de Puno a 4332 m.s.n.m. cuyas coordenadas geográficas son Longitud 70.34400114 y Latitud -16.140449

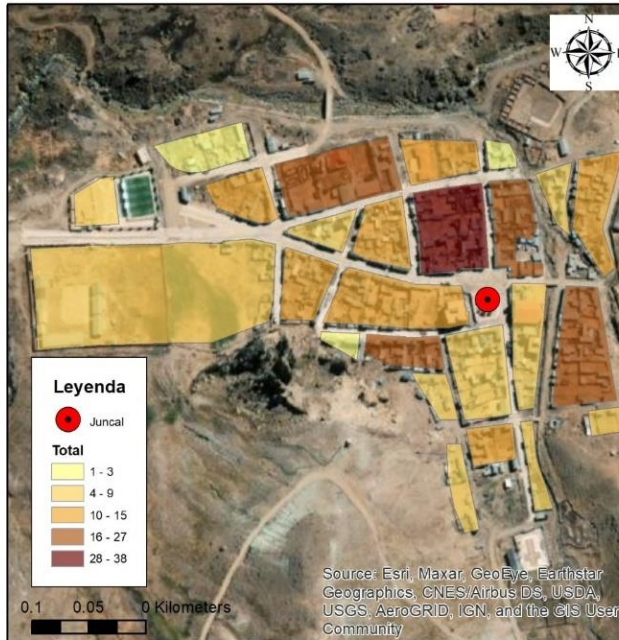


Fig. 1. Ubicación de la zona estudio y densidad poblacional a nivel de manzanas.

3.2 Toma de datos

El centro poblado el Juncal cuenta con una población de 288 personas distribuidas en 25 manzanas tal y como se muestra en la Fig. [31]. De acuerdo con la guía para la caracterización de RS municipales del ministerio del ambiente [32], para poblaciones de hasta 500 habitantes se recomienda tomar muestras de la generación de residuos de 54 familias, por lo que se tomaron los datos de 50 familias, las cuales se repartieron de manera proporcional a la densidad poblacional de las manzanas mostradas en la Fig. . Se recogió los datos socioeconómicos y demográficos de las familias mediante encuestas y se realizó el pesado de los residuos generados durante el día, el periodo de muestreo fue de 8 días anulando el primer día para con lo que se hizo el pesado completo de los residuos durante un día entero para una semana.

Con respecto a las variables socioeconómicas demográficas recogidas mediante encuestas se consideraron datos de ingreso económico [20], [21], [25], [33], el gasto [20], el tamaño de familia [20], [25]–[27], [33], servicios [25]–[27], educación [28], [29], la edad [30] y el empleo [28], [30]. También se consideró una variable adicional que refleja el nivel de educación ambiental en cuanto a RS y una variable que refleja el nivel de servicio contabilizando todos los servicios en la vivienda.

3.3 Análisis de correlación y significancia de las variables socioeconómicas demográficas en la generación per cápita RS

Los datos recogidos fueron de tipo categórico y numérico por lo que se realizaron 2 procedimientos para el análisis de correlación y significancia: con respecto a las variables numéricas se realizó la correlación de Spearman (ρ), seleccionando sólo las variables cuyo coeficiente de correlación fue mayor a 0,4 [13] y con respecto a las variables categóricas se realizó una prueba de significancia ANOVA (Análisis de Varianza), seleccionando sólo las variables cuya influencia en la generación per cápita de residuos fue significativa un nivel de significancia de $\alpha = 0,05$ [14].

3.4 Entrenamiento, validación del modelo y métricas de rendimiento

Una vez seleccionados los datos en el procedimiento anterior, se procedió a realizar el entrenamiento de los siguientes modelos: regresión lineal múltiple, RF, SVM, procesos gaussianos y redes neuronales. Para la validación de los primeros cuatro modelos se utilizó validación cruzada o k-fold, con el parámetro $k = 5$. La validación cruzada es un procedimiento de remuestreo que se utiliza para evaluar modelos de aprendizaje automático en una muestra de datos limitada [34], lo que corresponde al tamaño de la data de este estudio (50 puntos de análisis). Para el modelo de redes neuronales se aplicó la validación holdout, es decir validación por partición utilizando el 60% de los datos para el entrenamiento y el 40% de los datos para la validación. Todos los modelos mencionados anteriormente poseen parámetros y métodos para integrar la validación con el entrenamiento de los modelos, esta información se detalla en el siguiente link https://github.com/FrancoAlberto/GPC_ML.git.

Finalmente, las métricas de rendimiento utilizadas fueron el coeficiente de determinación R^2 , error cuadrático medio RMSE y el error absoluto medio MAE. En la Fig. 2, se muestra un flujograma del procedimiento.

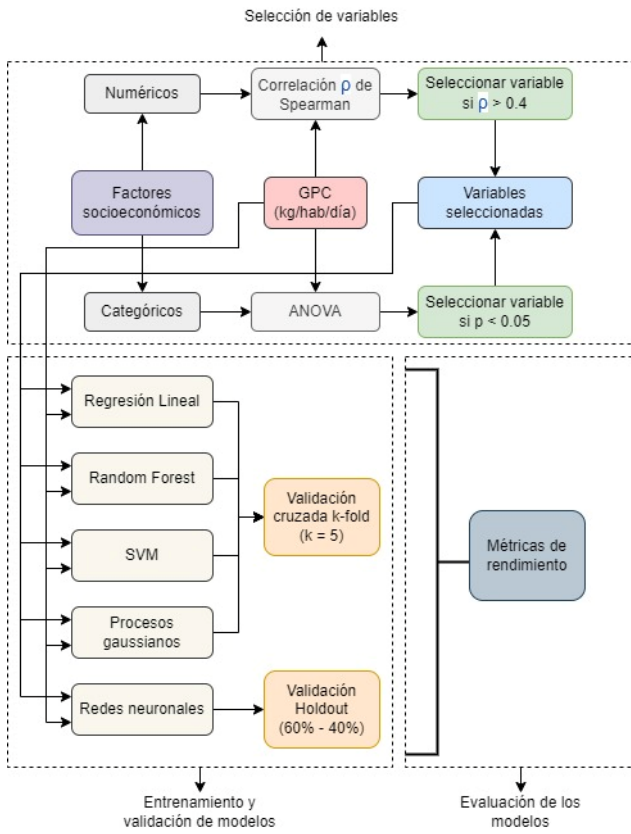


Fig. 2. Flujo de trabajo del proceso metodológico para la elaboración de modelo de Machine Learning

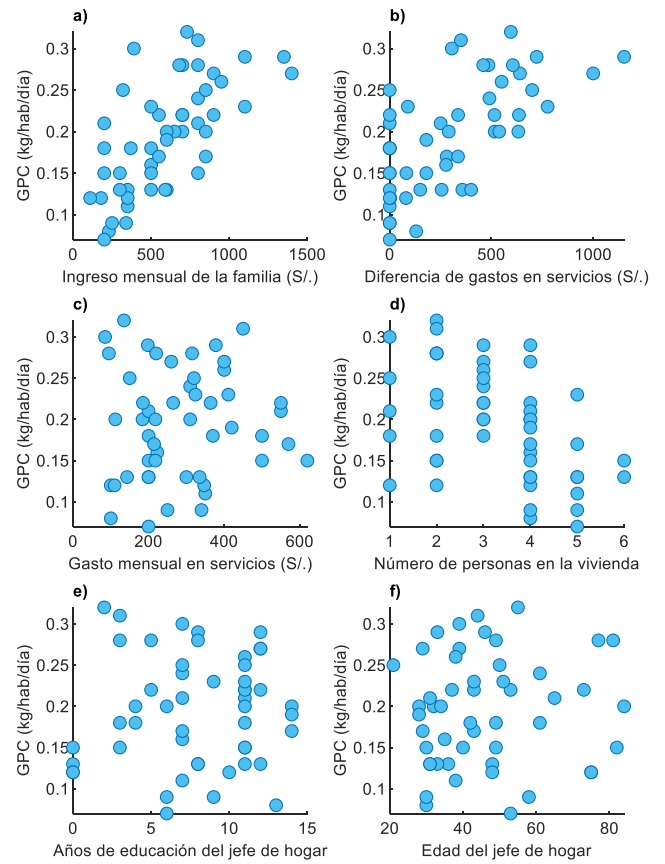


Fig. 3. Gráfico de dispersión de las variables numéricas socioeconómicas demográficas versus la generación per cápita de RS

4. ANÁLISIS DE RESULTADOS

4.1 Análisis de correlación y significancia de las variables socioeconómicas demográficas en la generación per cápita RS

De acuerdo con la Fig. 3. y la Tabla , en las figuras a), b), c), e) y f), existe una correlación positiva entre la generación per cápita y el ingreso mensual de la familia ($\rho = 0.68$), la diferencia de gastos en servicios, es decir el dinero que queda después de realizar los pagos de los servicios ($\rho = 0.65$), la edad del jefe de hogar ($\rho = 0.10$), los años de educación del jefe de hogar ($\rho = 0.07$) y el gasto mensual en servicios ($\rho = 0.06$). respectivamente.

Para la figura b) existe una correlación negativa entre la generación per cápita y el número de personas que habitan en la vivienda encuestada ($\rho = -0.49$).

De acuerdo con la Fig. 4. y la Tabla , en las figuras a), d), y h), existe evidencia estadística para afirmar que existe una influencia altamente significativa ($p - \text{valor} < 0.001$) de las variables socioeconómicas: servicio de agua, televisión en casa y cantidad de servicios sobre la generación per cápita de RS, en las figuras b) y c), se aprecia que existe evidencia estadística para afirmar que existe una influencia significativa ($p - \text{valor} < 0.05$) de las variables socioeconómicas: electricidad en las viviendas y celular en la vivienda sobre la generación per cápita de RS; y finalmente en las figuras e), f), y g) no se encuentra evidencia estadística ($p - \text{valor} > 0.05$) para afirmar que las variables genero del jefe del hogar, ocupación del jefe del hogar y si recibió o no alguna capacitación sobre los RS influyen significativamente sobre la generación per cápita de residuos sólido

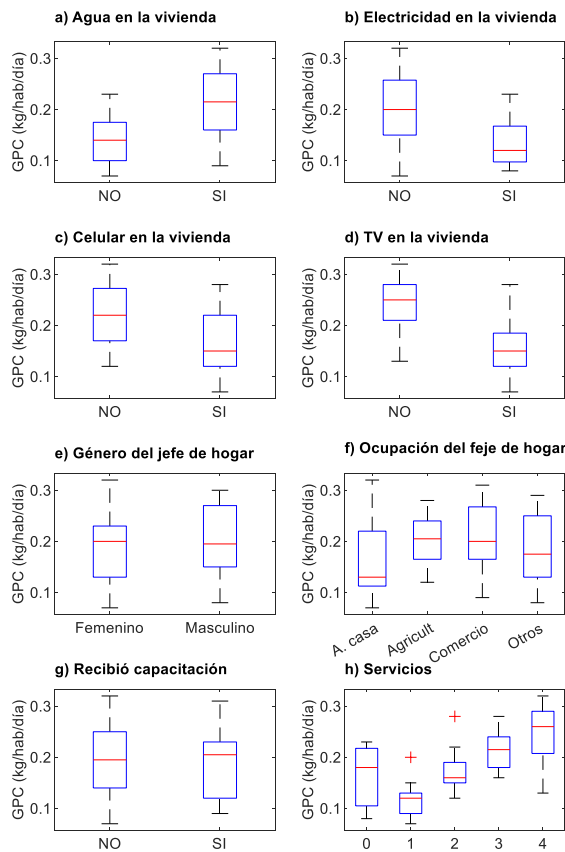


Fig. 4. Diagrama de cajas de las variables categóricas socioeconómicas demográficas con respecto a la generación per cápita de RS

Tabla I
Coeficientes de correlación ρ de Spearman para las variables numéricas y p – valor del ANOVA para las variables categóricas con las decisiones de selección o rechazo a partir de los umbrales planteados

Variable numérica		
Nombre	Coefficiente ρ	Decisión
Ingreso mensual de la familia (S./.)	0.68	Aceptado
Diferencia de gastos en servicios (S./.)	0.65	Aceptado
Gasto mensual en servicios (S./.)	0.06	Rechazado
Número de personas en la vivienda	-0.48	Aceptado
Años de educación del jefe de hogar	0.07	Rechazado
Edad del jefe de hogar	0.1	Rechazado
Variables categóricas		
Nombre	p – valor	Decisión
Agua en la vivienda	<0.001	Aceptado
Electricidad en la vivienda	0.01	Aceptado
Celular en la vivienda	0.004	Aceptado
TV en la vivienda	<0.001	Aceptado
Género del jefe de hogar	0.529	Rechazado
Ocupación del jefe de hogar	0.297	Rechazado
Recibió capacitación	0.961	Rechazado
Servicios	<0.001	Aceptado

La variable numérica mejor correlacionada con la generación per cápita de RS es la del ingreso mensual en soles de la familia ($\rho = 0.680$), lo que concuerda con el estudio de Vieira & Matheus [21], en donde se

encontró una correlación positiva entre el ingreso económico y la generación per cápita de residuos ($\rho = 0.607$), de acuerdo con Kumar et al., (2018) y Lebersorger & Beigl, (2011) [30], [35], es de esperarse que la cantidad de residuos generados aumenten de manera proporcional con los ingresos económicos ya que a mayor ingreso económico, mayor poder adquisitivo y como consecuencia mayor cantidad de residuos de las adquisiciones. La segunda variable numérica mejor correlacionada es la diferencia de gastos es decir la cantidad de dinero que queda disponible después de haber gastado en servicios como luz, agua, celular y/o televisión ($\rho = 0.650$), esta variable refleja la cantidad de dinero que se invierte en gastos de comida y productos del hogar, lo que significa una mayor proporción de RS orgánicos, el centro poblado el Juncal es una zona rural, se demuestra que a mayor ingresos mayor cantidad de RS y especialmente mayor cantidad de residuos orgánicos domiciliarios [21]. Con respecto a la variable de la cantidad de personas en el domicilio esto se correlaciona negativamente con la generación per cápita ($\rho = -0.480$), esto es consistente con los trabajos de Cha et al., (2017) y Dai et al., (2011) [17], [26]. De acuerdo con los autores citados la cantidad de residuos per cápita en el hogar disminuye con la cantidad de personas debido a los productos comunes que se utilizan en una vivienda, por ejemplo, insumos de limpieza, alimentos y otros se consumen en cantidades diferentes, pero generan casi la misma cantidad de residuos. La correlación entre los gastos en servicios del hogar y la generación per cápita es casi nula ($\rho = 0.06$), esto se explica porque los gastos en los servicios no generan los residuos en las viviendas, los residuos se generan por la compra de insumos y productos necesarios para el hogar, la correlación entre los años acumulados de estudio del jefe de hogar y la generación per cápita también es nula ($\rho = 0.07$), esto difiere de las investigaciones de Thanh et al., (2010), Camero et al., (2019) y Soni et al., (2019) [4], [8], [27]; de acuerdo con Vieira & Matheus, (2018) [21] aunque el grado de instrucción sí influye en la cantidad de residuos que generan las viviendas esta influencia decrece a medida que las viviendas se alejan de las ciudades y son más cercanas a zonas rurales, en el caso de nuestra investigación el centro poblado el Juncal es una zona rural por lo que podría explicarse la nula correlación encontrada.

Con respecto a los servicios del hogar (agua, electricidad TV y celular), se encontró influencia significativa en la generación per cápita de residuos sólido. De acuerdo con Niska & Serkkola, (2018) [19] en donde se identificaron perfiles de generación de RS, se identificó que mayores ingresos económicos están asociados con mayor cantidad de servicios en la vivienda, por lo que la influencia significativa de los servicios sobre la generación per cápita de RS, más que una causalidad puede estar reflejando el nivel

adquisitivo de las personas. la influencia de las capacitaciones sobre la generación de RS no fue significativa, como explica Vieira & Matheus, (2018) [21], las medidas de educación ambiental no son efectiva si es que no están acompañadas de medidas estructurales, como es el caso del centro poblado el Juncal.

Aplicación y validación de modelos de Machine Learning

Como se puede apreciar en la **Tabla** y la **Fig. 5** el modelo que tuvo mejor desempeño fue el de SVM, con un $R^2 = 0.986$, el promedio del error absoluto de predicción de generación de RS es de $ECM = 0.007$ kg o 7 g; y la raíz cuadrada del error cuadrático medio en la predicción de generación de RS es de $RECM = 0.008$ kg o 8 g: El fallo promedio del modelo de SVM es de 7 a 8 g en peso.

Tabla II

Coefficiente de desempeño (error absoluto medio EAM, la raíz del error cuadrático medio RECM y el coeficiente de determinación R^2) de los modelos de machine learning para la predicción de la generación de RS a partir de variables socioeconómicas demográficas

Modelos	EAM	RECM	R^2
Regresión	0.011	0.013	0.959
Random Forest	0.022	0.027	0.832
Support Vector Machine	0.007	0.008	0.986
Procesos Gaussianos	0.007	0.009	0.982
Redes Neuronales	0.011	0.016	0.942

El modelo que tuvo peor desempeño fue el RF, con un $R^2 = 0.832$, el promedio del error absoluto de predicción de generación de RS es de $ECM = 0.022$ kg o 22 g; y la raíz cuadrada del error cuadrático medio en la predicción de generación de RS es de $RECM = 0.027$ kg o 27 g: El fallo promedio del modelo RF es de 22 a 27 g en peso.

El modelo de Machine learning que tuvo mejor desempeño de predicción fue el modelo de SVM, de acuerdo con Solano Meza et al., (2019) [6] este modelo se ajusta bien a los datos de entrenamiento de RS a pesar de su pequeño tamaño, con respecto al coeficiente de determinación, el valor es semejante al encontrado en el estudio de Abdallah et al., (2020) [12], con $R^2 = 0.932$ en comparación al de este estudio $R^2 = 0.986$ en comparación al de este estudio $R^2 = 0.986$ El segundo modelo con mejor desempeño fue el de procesos gaussianos con un coeficiente de determinación de $R^2 = 0.982$, este valor se asemeja al obtenido en el estudio de Ceylan, (2020) [15] quien aplicando también un modelo de procesos gaussianos en la estimación de la generación de RS, encontró un coeficiente de determinación de $R^2 = 0.991$.

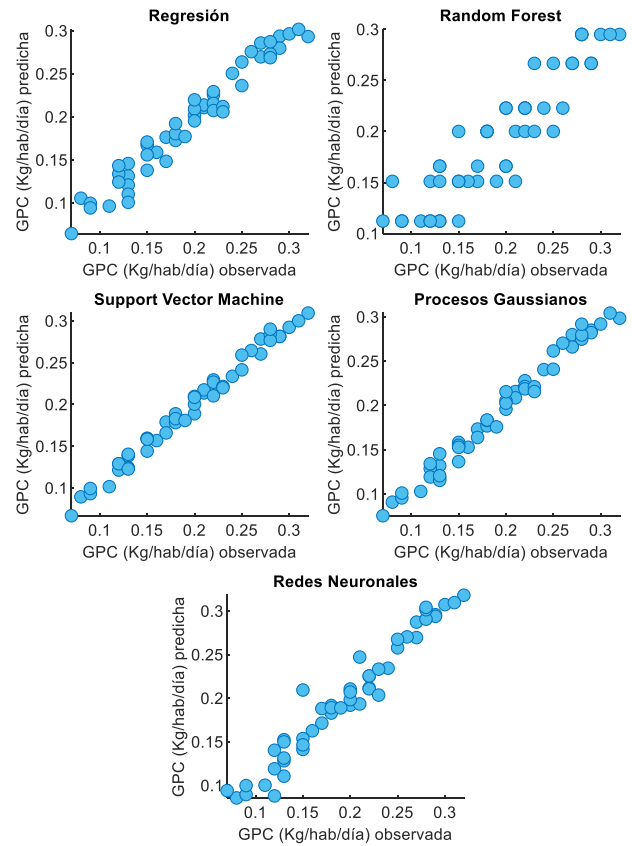


Fig. 5. Comparación gráfica del desempeño de los modelos de machine learning para la predicción de la generación de residuos a partir de variables socioeconómicas y demográfica.

Con respecto al modelo de regresión lineal múltiple, el coeficiente de determinación obtenido fue de $R^2 = 0.959$, valor que se asemeja al obtenido en el estudio de Golbaz et al., (2019) [10], con un $R^2 = 0.96$. Los resultados obtenidos con los modelos de regresión son fáciles de interpretar y los costes computacionales son bajos, sin embargo, este método se considera generalmente inadecuado para modelar datos no lineales [15]. El modelo de redes neuronales obtuvo un coeficiente de determinación de $R^2 = 0.942$, siendo uno de los menores rendimientos obtenidos. Según Abdallah et al., (2020) [12], aunque las redes neuronales se han aplicado exitosamente a numerosas situaciones, incluida la gestión de RS, estas, son débiles en el manejo de problemas lógicos y aritméticos que requieren una gran precisión y son propensas al sobreajuste. Además, las RNA son incapaces de determinar la importancia relativa de los de los numerosos factores que intervienen en el análisis, es decir, qué característica de entrada tiene el mayor impacto en la salida, Como en el caso de este estudio en las que unas variables serán más importantes que otras. Finalmente, el modelo de menor desempeño fue el de RF ($R^2 = 0.832$), Según Abdallah et al., (2020) [12] una de las dificultades del modelo de RF en cuanto a las regresiones es que tiende a discretizar demasiado los valores continuos, y tiene mejor desempeño en los problemas de clasificación.

Todos estos modelos tienen desempeños desde aceptables hasta muy buenos, y tienen como insumos variables socioeconómicas y demográficas que actualmente se encuentran disponibles en bases de datos abiertas como son el REDATAM del INEI, y a partir de esta información se pueden estimar las cantidades de RS que se podrían generar en zonas urbanas y rurales en las que no se cuenta con estudios y poder a partir de esta información plantear alternativas de gestión

CONCLUSIONES

- Las variables socioeconómicas demográficas numéricas que tuvieron influencia en la generación de RS en el centro poblado el Juncal fueron el ingreso familiar ($\rho = 0.68$), el gasto en productos e insumos (no servicios) ($\rho = 0.65$) y la cantidad de personas en las viviendas ($\rho = -0.48$). Las variables categóricas influyentes ($p - \text{valor} < 0.05$) son los servicios en las viviendas (agua electricidad TV y celulares) que más que reflejar una causalidad refuerza la influencia del ingreso económico en la generación de residuos.
- Los modelos de Machine learning tuvieron un desempeño aceptable con coeficiente de determinación de $R^2 = 0.986, 0.982, 0.959, 0.942, 0.832$; para los modelos de SVM, procesos gaussianos, regresión lineal múltiple, redes neuronales y RF respectivamente. La raíz cuadrada del error cuadrático medio en los modelos es desde 8 gramos hasta 27 gramos de error de predicción, estos valores son relativamente pequeños por lo que estos modelos pueden ser utilizados para la estimación de la generación de RS per cápita a partir de factores socioeconómicos y demográficos y esta información puede utilizarse para el diseño de proyectos

REFERENCIAS

- [1] INEI, "Cantidad promedio diaria de residuos sólidos (basura) recolectada, según departamento, 2018", *Compendio Estadístico 2020*, 2020.
- [2] DGGRS, "Valorización de residuos sólidos orgánicos municipales", 2018.
- [3] MINAM, "Portal Web SINIA: Residuos", *Sistema Nacional de Información Ambiental*, 2021. <https://sinia.minam.gob.pe/informacion/tematicas?tematica=08> (consultado ago. 04, 2021).
- [4] U. Soni, A. Roy, A. Verma, y V. Jain, "Forecasting municipal solid waste generation using artificial intelligence models—a case study in India", *SN Appl. Sci.*, vol. 1, no. 2, pp. 1–10, 2019, doi: 10.1007/s42452-018-0157-x.
- [5] A. P. Condori Iquise, "Factores socioeconómicos que inciden en la producción de residuos sólidos en el distrito de San Antonio de Esquilache, año 2015", Universidad Nacional del Altiplano, 2017.
- [6] J. K. Solano Meza, D. Orjuela Yepes, J. Rodrigo-llarri, y E. Cassiraga, "Predictive analysis of urban waste generation for the city of Bogotá, Colombia, through the implementation of decision trees-based machine learning, support vector machines and artificial neural networks", *Heliyon*, vol. 5, no. 11, p. e02810, 2019, doi: 10.1016/j.heliyon.2019.e02810.
- [7] N. E. Johnson et al., "Patterns of waste generation: A gradient boosting model for short-term waste prediction in New York City", *Waste Manag.*, vol. 62, pp. 3–11, 2017, doi: 10.1016/j.wasman.2017.01.037.
- [8] A. Camero, J. Toutouh, J. Ferrer, y E. Alba, "Waste Generation Prediction in Smart Cities Through Deep Neuroevolution", *Commun. Comput. Inf. Sci.*, vol. 978, pp. 192–204, 2019, doi: 10.1007/978-3-030-12804-3_15.
- [9] A. Kumar, S. R. Samadder, N. Kumar, y C. Singh, "Estimation of the generation rate of different types of plastic wastes and possible revenue recovery from informal recycling", *Waste Manag.*, vol. 79, pp. 781–790, 2018, doi: 10.1016/j.wasman.2018.08.045.
- [10] S. Golbaz, R. Nabizadeh, y H. S. Sajadi, "Comparative study of predicting hospital solid waste generation using multiple linear regression and artificial intelligence", *J. Environ. Heal. Sci. Eng.*, vol. 17, no. 1, pp. 41–51, 2019, doi: 10.1007/s40201-018-00324-z.
- [11] C. E. Kontokosta, B. Hong, N. E. Johnson, y D. Starobin, "Using machine learning and small area estimation to predict building-level municipal solid waste generation in cities", *Comput. Environ. Urban Syst.*, vol. 70, no. March, pp. 151–162, 2018, doi: 10.1016/j.compenvurbsys.2018.03.004.
- [12] M. Abdallah, M. Abu Talib, S. Feroz, Q. Nasir, H. Abdalla, y B. Mahfood, "Artificial intelligence applications in solid waste management: A systematic research review", *Waste Manag.*, vol. 109, pp. 231–246, 2020, doi: 10.1016/j.wasman.2020.04.057.
- [13] M. Kannagara, R. Dua, L. Ahmadi, y F. Bensebaa, "Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches", *Waste Manag.*, vol. 74, pp. 3–15, 2018, doi: 10.1016/j.wasman.2017.11.057.
- [14] V. M. Adamović, D. Z. Antanasijević, M. Ristić, A. A. Perić-Grujić, y V. V. Pocajt, "An optimized artificial neural network model for the prediction of rate of hazardous chemical and healthcare waste generation at the national level", *J. Mater. Cycles Waste Manag.*, vol. 20, no. 3, pp. 1736–1750, 2018, doi: 10.1007/s10163-018-0741-6.
- [15] Z. Ceylan, "Estimation of municipal waste

- generation of Turkey using socio-economic indicators by Bayesian optimization tuned Gaussian process regression”, *Waste Manag. Res.*, vol. 38, no. 8, pp. 840–850, 2020, doi: 10.1177/0734242X20906877.
- [16] L. Chhay, M. A. H. Reyad, R. Suy, M. R. Islam, y M. M. Mian, “Municipal solid waste generation in China: influencing factor analysis and multi-model forecasting”, *J. Mater. Cycles Waste Manag.*, vol. 20, no. 3, pp. 1761–1770, 2018, doi: 10.1007/s10163-018-0743-4.
- [17] G. W. Cha, Y. C. Kim, H. J. Moon, y W. H. Hong, “New approach for forecasting demolition waste generation using chi-squared automatic interaction detection (CHAID) method”, *J. Clean. Prod.*, vol. 168, pp. 375–385, 2017, doi: 10.1016/j.jclepro.2017.09.025.
- [18] F. Wu, D. Niu, S. Dai, y B. Wu, “New insights into regional differences of the predictions of municipal solid waste generation rates using artificial neural networks”, *Waste Manag.*, vol. 107, pp. 182–190, 2020, doi: 10.1016/j.wasman.2020.04.015.
- [19] H. Niska y A. Serkkola, “Data analytics approach to create waste generation profiles for waste management and collection”, *Waste Manag.*, vol. 77, pp. 477–485, 2018, doi: 10.1016/j.wasman.2018.04.033.
- [20] R. Intharathirat, P. Abdul Salam, S. Kumar, y A. Untong, “Forecasting of municipal solid waste quantity in a developing country using multivariate grey models”, *Waste Manag.*, vol. 39, pp. 3–14, may 2015, doi: 10.1016/j.wasman.2015.01.026.
- [21] V. H. A. de M. Vieira y D. R. Matheus, “The impact of socioeconomic factors on municipal solid waste generation in São Paulo, Brazil”, *Waste Manag. Res.*, vol. 36, no. 1, pp. 79–85, 2018, doi: 10.1177/0734242X17744039.
- [22] T. V. Ramachandra, H. A. Bharath, G. Kulkarni, y S. S. Han, “Municipal solid waste: Generation, composition and GHG emissions in Bangalore, India”, *Renew. Sustain. Energy Rev.*, vol. 82, núm. September 2017, pp. 1122–1136, 2018, doi: 10.1016/j.rser.2017.09.085.
- [23] K. A. Kolekar, T. Hazra, y S. N. Chakrabarty, “A Review on Prediction of Municipal Solid Waste Generation Models”, *Procedia Environ. Sci.*, vol. 35, pp. 238–244, 2016, doi: 10.1016/j.proenv.2016.07.087.
- [24] P. Beigl, S. Lebersorger, y S. Salhofer, “Modelling municipal solid waste generation: A review”, *Waste Manag.*, vol. 28, no. 1, pp. 200–214, 2008, doi: 10.1016/j.wasman.2006.12.011.
- [25] S. S. Chung, “Projecting municipal solid waste: The case of Hong Kong SAR”, *Resour. Conserv. Recycl.*, vol. 54, no. 11, pp. 759–768, 2010, doi: 10.1016/j.resconrec.2009.11.012.
- [26] C. Dai, Y. P. Li, y G. H. Huang, “A two-stage support-vector-regression optimization model for municipal solid waste management - A case study of Beijing, China”, *J. Environ. Manage.*, vol. 92, no. 12, pp. 3023–3037, 2011, doi: 10.1016/j.jenvman.2011.06.038.
- [27] N. P. Thanh, Y. Matsui, y T. Fujiwara, “Household solid waste generation and characteristic in a Mekong Delta city, Vietnam”, *J. Environ. Manage.*, vol. 91, no. 11, pp. 2307–2321, 2010, doi: 10.1016/j.jenvman.2010.06.016.
- [28] S. Keser, S. Duzgun, y A. Aksoy, “Application of spatial and non-spatial data analysis in determination of the factors that impact municipal solid waste generation rates in Turkey”, *Waste Manag.*, vol. 32, no. 3, pp. 359–371, 2012, doi: 10.1016/j.wasman.2011.10.017.
- [29] S. O. Benítez, G. Lozano-Olvera, R. A. Morelos, y C. A. de Vega, “Mathematical modeling to predict residential solid waste generation”, *Waste Manag.*, vol. 28, no. SUPPL. 1, pp. 7–13, 2008, doi: 10.1016/j.wasman.2008.03.020.
- [30] S. Lebersorger y P. Beigl, “Municipal solid waste generation in municipalities: Quantifying impacts of household structure, commercial waste and domestic fuel”, *Waste Manag.*, vol. 31, no. 9–10, pp. 1907–1915, 2011, doi: 10.1016/j.wasman.2011.05.016.
- [31] INEI, “INEI pone a disposición del país dos sistemas de consulta sobre las características de la población y vivienda a nivel de manzana”, 2017. <https://www.inei.gob.pe/prensa/noticias/inei-pone-a-disposicion-del-pais-dos-sistemas-de-consulta-sobre-las-caracteristicas-de-la-poblacion-y-vivienda-a-nivel-de-manzana-12162/> (consultado ago. 04, 2021).
- [32] MINAM, “Guía para la caracterización de residuos sólidos municipales”, 2019. [En línea]. Disponible en: https://cdn.www.gob.pe/uploads/document/file/523785/Guía_para_la_caracterización_rsm-29012020__1_.pdf.
- [33] T. V. Ramachandra, H. A. Bharath, G. Kulkarni, y S. S. Han, “Municipal solid waste: Generation, composition and GHG emissions in Bangalore, India”, *Renew. Sustain. Energy Rev.*, vol. 82, no. September 2017, pp. 1122–1136, 2018, doi: 10.1016/j.rser.2017.09.085.
- [34] I. Kononenko y M. Kukar, *Machine learning and data mining: introduction to principles and algorithms*, 2da ed., vol. 45, no. 07, West Sussex: Horwood Publishing Chichester, 2007.
- [35] A. Kumar, S. R. Samadder, N. Kumar, y C. Singh, “Estimation of the generation rate of different types of plastic wastes and possible revenue recovery from informal recycling”, *Waste Manag.*, vol. 79, pp. 781–790, 2018, doi:

10.1016/j.wasman.2018.08.045.



Los artículos publicados por TECNIA pueden ser compartidos a través de la licencia Creative Commons: CC BY 4.0. Permisos lejos de este alcance pueden ser consultados a través del correo revistas@uni.edu.pe