

**PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ**

Escuela de Posgrado



**Propuesta metodológica para la optimización de modelos
predictivos de generación de residuos sólidos municipales
en zonas urbanas**

Tesis para obtener el grado académico de Doctor en Ingeniería que
presenta:

Luis Antonio Izquierdo Horna

Asesor: ***Ramzy Francis Kahhat Abedrabbo***

Co-Asesor: ***Ian Vázquez Rowe***


Lima, 2024

INFORME DE SIMILITUD

Yo, **Ramzy Francis Kahhat Abedrabbo**, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis titulada **“Propuesta metodológica para la optimización de modelos predictivos de generación de residuos sólidos municipales en zonas urbanas”**, del autor Luis Antonio Izquierdo Horna, dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 18 %. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 26/05/2024.
- He revisado con detalle dicho reporte y la Tesis y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lima, 26 de mayo - 2024.

Apellidos y nombres del asesor: Kahhat Abedrabbo, Ramzy Francis	
DNI : 08746262	Firma: 
ORCID : 0000-0001-7321-2256	

AGRADECIMIENTOS

El autor desea expresar su gratitud a sus asesores, Dr. Ramzy Kahhat y Dr. Ian Vázquez por toda la guía y supervisión brindada durante este período de formación. Sus consejos han sido invaluable en el desarrollo de este trabajo de investigación. Asimismo, desea agradecer a todas las personas y organizaciones que contribuyeron a la realización de este proyecto.

RESUMEN

El pronóstico de la generación de residuos sólidos municipales (RSM) desempeña un papel esencial en la toma de decisiones y proporciona información relevante para la gestión de residuos, así como una comprensión profunda de los factores que influyen en este proceso. En este trabajo, se desarrolló un modelo de predicción de RSM específico para Lima Metropolitana, basado en variables socioculturales, ambientales y económicas, teniendo al 2019 como año de referencia, debido a la influencia del COVID-19 en los datos sobre este tema en años posteriores a la pandemia. El modelo se construyó utilizando las cantidades per cápita de RSM generadas en cada distrito, junto con parámetros relacionados con el consumo de combustibles en el hogar (como gas natural, electricidad y gas licuado de petróleo) y características demográficas de la población (como edad, nivel de educación y gasto mensual). Dada la calidad y disponibilidad de datos, se optó por utilizar el algoritmo de *random forest* como técnica de predicción. Las variables analizadas se obtuvieron a partir de la Encuesta Residencial de Consumo y Uso de Energía (ERCUE) a nivel municipal. Los resultados indicaron que el algoritmo implementado explica el 51% de la variabilidad de los datos. Se espera que las recomendaciones presentadas en este estudio sirvan para investigaciones futuras relacionadas con la predicción de RSM, contribuyendo a obtener resultados más precisos y aplicables a contextos específicos.

Palabras clave: Aprendizaje automático, Consumo de combustible, Predicción, *Random forest*, Residuos sólidos municipales, Perú

ABSTRACT

Municipal solid waste (MSW) generation forecasting plays an essential role in decision making and provides relevant information for waste management, as well as a deep understanding of the factors that influence this process. In this work, a specific MSW prediction model was developed for Metropolitan Lima, based on sociocultural, environmental and economic variables, having 2019 as the reference year, due to the influence of COVID-19 on data on this topic in post-pandemic years. The model was constructed using per capita amounts of MSW generated in each district, along with parameters related to household fuel consumption (such as natural gas, electricity, and liquefied petroleum gas) and demographic characteristics of the population (such as age, education level, and monthly expenditure). Given the quality and availability of data, we chose to use the random forest algorithm as a prediction technique. The variables analyzed were obtained from the Residential Survey of Energy Consumption and Use (ERCUE) at the municipal level. The results indicated that the implemented algorithm explains 51% of the variability of the data. It is expected that the recommendations presented in this study will be useful for future research related to MSW prediction, contributing to obtain more accurate results applicable to specific contexts.

Keywords: Fuel consumption, Forecasting, Machine learning, Municipal solid waste, Random forest, Peru

RESUMEN EJECUTIVO

La gestión de residuos sólidos municipales (RSM) se ha convertido en un desafío cada vez más apremiante a nivel mundial debido al crecimiento acelerado de la población, la expansión urbana y el aumento del consumo. Esta situación plantea retos tanto en términos de infraestructura para la gestión de RSM como en la implementación de políticas efectivas de reducción, reutilización y reciclaje. En este contexto, la investigación y la innovación desempeñan un papel crucial en el desarrollo de enfoques más efectivos y sostenibles para la gestión de RSM. Desde el análisis de patrones de generación hasta la evaluación de tecnologías de tratamiento y la implementación de políticas públicas, existe una amplia gama de áreas de estudio que pueden contribuir a abordar esta problemática de manera integral y multidisciplinaria. De esta manera, entender los factores subyacentes que inciden en la generación de residuos sólidos se ha vuelto una prioridad fundamental para investigadores, formuladores de políticas y profesionales del sector. Esta necesidad cobra aún más relevancia al considerar que la composición y las características de los residuos sólidos varían significativamente según la región, la cultura y los hábitos de consumo. Esta diversidad añade una complejidad adicional a la gestión de residuos, destacando la importancia de desarrollar modelos predictivos precisos que puedan adaptarse a las particularidades de cada contexto.

Así pues, la identificación oportuna de las variables que influyen en la generación de RSM permite generar el escenario para una gestión eficiente de RSM contemplando su variabilidad en el espacio y en el tiempo. Además, la disponibilidad y confiabilidad de los datos juegan un papel crítico tanto en la determinación de estas variables clave como en el establecimiento de los canales de acceso a la información. Algunos estudios utilizan censos, bases de datos proporcionadas por organizaciones relacionadas, encuestas o datos geográficos para recopilar datos primarios. En esta línea, se han identificado dos vacíos en la investigación a partir de la revisión bibliográfica. La primera es la falta de preprocesamiento para determinar e identificar

variables para caracterizar a la población, y la segunda es el grado de disponibilidad de los datos para la implementación de un modelo estadístico posterior.

El objetivo principal de esta investigación es implementar un modelo de predicción de RSM a nivel vivienda, aprovechando información de acceso libre, especialmente aquella relacionada con las características propias de cada hogar. Para lograr este propósito, se han delineado objetivos específicos. En primer lugar, analizar los patrones de consumo en los hogares urbanos, basándose en la literatura existente y en datos disponibles. Asimismo, identificar un conjunto de variables con potencial predictivo para la generación de RSM, mediante la caracterización de las condiciones de vida y el entorno construido. Finalmente, diseñar e implementar un algoritmo de *random forest* para la predicción de la generación de RSM a nivel vivienda, utilizando datos de acceso público como encuestas gubernamentales.

En cuanto al alcance de la investigación, esta se centra exclusivamente en el análisis de los aspectos relacionados con la generación de RSM en entornos urbanos, específicamente en Lima Metropolitana. Se utilizarán datos del año 2019, considerando las complicaciones causadas por la pandemia de la COVID-19. Se espera que la metodología empleada sea replicable y adaptable a diferentes casos de estudio, sujeta a la validación de las condiciones específicas de cada contexto.

En consecuencia, se realizó una revisión semi-sistemática cuyo objetivo primario fue examinar la influencia de las variables predictoras en el proceso de generación de RSM. Esta revisión reconoce la interrelación entre factores socioculturales, económicos y ambientales, los cuales influyen significativamente en la generación de residuos. Se destaca que la elección de estas variables predictoras no solo se basa en su capacidad predictiva, sino también en la disponibilidad de datos, lo que resalta la importancia de considerar múltiples dimensiones al desarrollar modelos de predicción. Esta revisión concluye proponiendo recomendaciones sobre

desafíos actuales e investigaciones futuras, enfatizando la importancia de abordar los aspectos socioculturales, económicos y ambientales en futuros estudios para mejorar la gestión de residuos sólidos de manera efectiva y sostenible. Entre las variables recomendadas se incluyen la edad, las condiciones demográficas, el nivel de educación, las características culturales, así como variables ambientales como la humedad, la velocidad del viento, la precipitación y la temperatura, y variables económicas como el ingreso económico.

Por otro lado, la ausencia de información primaria confiable y actualizada sobre la generación de RSM hizo necesaria la inclusión de *proxies* asociados al consumo de combustible para comprender el proceso de generación de residuos sólidos domiciliarios y su incidencia en la fracción orgánica, específicamente con la generación de residuos alimenticios. Para validar esta relación, se implementó un modelo de regresión lineal múltiple contemplando las variables disponibles, centrándose en el consumo del gas licuado de petróleo (GLP), gas natural (GN) y electricidad en suelos residenciales. Los resultados sugieren que el consumo combinado de GLP & GN y el de electricidad guardan significancia con la generación de residuos alimenticios domiciliarios. La correlación encontrada entre la generación de residuos de alimentos y estas variables indica el potencial para desarrollar un modelo de pronóstico de generación de residuos, aspecto relevante dada la limitada disponibilidad de datos en la gestión de residuos.

Seguidamente, se implementó un modelo de predicción basado en el consumo de combustible para cocinar, las características del hogar obtenidas del ERCUE - OSINERGMIN y la generación per cápita a nivel distrital extraída del SIGERSOL - MINAM. Dado que se contaba con una cantidad limitada de datos de calidad, se optó por el algoritmo random forest. Para evaluar el desempeño del modelo, se utilizaron métricas como el coeficiente R² (coeficiente de determinación) y el MSE (error cuadrático medio). El coeficiente R² indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables

independientes, mientras que el MSE mide la calidad de la predicción al cuantificar la diferencia cuadrática promedio entre los valores predichos y los observados. Los resultados del entrenamiento revelaron que el modelo logra capturar el 51 % de la variabilidad presente en los datos originales. Sin embargo, es importante destacar que, dada la complejidad y la diversidad de factores que influyen en la generación de residuos sólidos domiciliarios, existe margen para mejorar la precisión del modelo mediante la inclusión de variables adicionales y la exploración de técnicas de modelado más avanzadas.

En conclusión, esta investigación subraya la necesidad de desarrollar metodologías sólidas para comprender y predecir la generación de RSM, especialmente en entornos urbanos donde la falta de datos primarios es común. Al superar desafíos como la identificación de variables relevantes y el uso de fuentes de información indirectas, hemos avanzado hacia una comprensión más profunda de los factores que influyen en la producción de residuos. Este enfoque multidisciplinario no solo proporciona una visión más clara de los patrones de consumo y comportamiento, sino que también abre nuevas vías para la gestión sostenible de los residuos, destacando la importancia de la colaboración entre investigadores, responsables políticos y comunidades locales para abordar este desafío de manera efectiva y con un resultado beneficioso para todos.

EXECUTIVE ABSTRACT

Municipal solid waste (MSW) management has increasingly become a pressing global challenge due to rapid population growth, urban expansion, and increased consumption. This situation poses challenges both in terms of infrastructure for MSW management and in the implementation of effective policies for reduction, reuse, and recycling. In this context, research and innovation play a crucial role in developing more effective and sustainable approaches to MSW management. From analyzing generation patterns to evaluating treatment technologies and implementing public policies, there is a wide range of areas of study that can contribute to addressing this issue comprehensively and multidisciplinary. Thus, understanding the underlying factors influencing solid waste generation has become a fundamental priority for researchers, policymakers, and industry professionals. This need is further underscored by the fact that the composition and characteristics of solid waste vary significantly depending on the region, culture, and consumption habits. This diversity adds an additional complexity to waste management, highlighting the importance of developing accurate predictive models that can adapt to the particularities of each context.

Therefore, the timely identification of variables influencing MSW generation sets the stage for efficient MSW management considering its variability in space and time. Moreover, the availability and reliability of data play a critical role in determining these key variables and establishing channels for accessing information. Some studies utilize censuses, databases provided by related organizations, surveys, or geographic data to collect primary data. In this line, two gaps in research have been identified through literature review. The first is the lack of preprocessing to determine and identify variables to characterize the population, and the second is the degree of data availability for the implementation of subsequent statistical modeling.

The main objective of this research is to implement a prediction model for MSW at the household level, leveraging freely accessible information, especially those related to the unique characteristics of each household. To achieve this purpose, specific objectives have been outlined. Firstly, to analyze consumption patterns in urban households, based on existing literature and available data. Additionally, to identify a set of variables with predictive potential for MSW generation by characterizing living conditions and the built environment. Finally, to design and implement a random forest algorithm for predicting MSW generation at the household level, using publicly accessible data such as government surveys.

Regarding the scope of the research, it focuses exclusively on analyzing aspects related to MSW generation in urban environments, specifically in Metropolitan Lima. Data from the year 2019 will be used, considering the complications caused by the COVID-19 pandemic. It is expected that the methodology employed will be replicable and adaptable to different case studies, subject to validation of the specific conditions of each context.

Consequently, a semi-systematic review was conducted whose primary objective was to examine the influence of predictor variables on the process of MSW generation. This review acknowledges the interrelation between sociocultural, economic, and environmental factors, which significantly influence waste generation. It is emphasized that the choice of these predictor variables is based not only on their predictive capacity but also on data availability, highlighting the importance of considering multiple dimensions when developing prediction models. This review concludes by proposing recommendations on current challenges and future research, emphasizing the importance of addressing sociocultural, economic, and environmental aspects in future studies to improve waste management effectively and sustainably. Recommended variables include age, demographic conditions, level of education, cultural characteristics, as well as environmental variables such as humidity, wind speed, precipitation, and temperature, and economic variables such as income.

On the other hand, the absence of reliable and up-to-date primary information on MSW generation made it necessary to include proxies associated with fuel consumption to understand the process of household solid waste generation and its impact on the organic fraction, specifically with food waste generation. To validate this relationship, a multiple linear regression model was implemented considering available variables, focusing on the consumption of liquefied petroleum gas (LPG), natural gas (NG), and electricity in residential areas. The results suggest that the combined consumption of LPG & NG and electricity is significant with the generation of household food waste. The correlation found between food waste generation and these variables indicates the potential to develop a waste generation forecasting model, a relevant aspect given the limited availability of data in waste management. Subsequently, a prediction model based on cooking fuel consumption, household characteristics obtained from ERCUE - OSINERGMIN, and per capita generation at the district level extracted from SIGERSOL - MINAM was implemented. Since there was a limited amount of quality data, the random forest algorithm was chosen. To evaluate the model's performance, metrics such as the R² coefficient (coefficient of determination) and MSE (mean squared error) were used. The R² coefficient indicates the proportion of variance in the dependent variable that is predictable from the independent variables, while MSE measures the prediction quality by quantifying the average squared difference between predicted and observed values. Training results revealed that the model captures 51% of the variability present in the original data. However, it is important to note that, given the complexity and diversity of factors influencing household solid waste generation, there is room to improve the model's accuracy by including additional variables and exploring more advanced modeling techniques.

In conclusion, this research underscores the need to develop robust methodologies to understand and predict MSW generation, especially in urban environments where the lack of

primary data is common. By overcoming challenges such as identifying relevant variables and using indirect sources of information, we have made progress toward a deeper understanding of the factors influencing waste production. This multidisciplinary approach not only provides a clearer insight into consumption and behavior patterns but also opens new avenues for sustainable waste management, highlighting the importance of collaboration among researchers, policymakers, and local communities to effectively address this challenge with beneficial outcomes for all.



TABLA DE CONTENIDOS

AGRADECIMIENTOS	i
RESUMEN	ii
ABSTRACT.....	iii
RESUMEN EJECUTIVO.....	iv
EXECUTIVE ABSTRACT	viii
LISTA DE TABLAS	xvii
LISTA DE FIGURAS.....	xix
CAPÍTULO 1: INTRODUCCIÓN	1
CAPÍTULO 2: PLANTEAMIENTO DE LA INVESTIGACIÓN	4
2.1 Diagnóstico del contexto actual	4
2.2 Abordando la complejidad en el pronóstico de RSM	5
2.3 Definición del problema de investigación.....	6
2.4 Objetivos de la investigación	9
2.5 Alcance, justificación y limitaciones	10
2.5.1. Alcance de la investigación	10
2.5.2. Justificación de la investigación	10
2.5.3. Limitaciones de la investigación.....	11
2.6 Estructura de la tesis y producción científica.....	12
CAPÍTULO 3: MARCO CONCEPTUAL DE LA INVESTIGACIÓN.....	14
3.1 Caracterización y valorización de los residuos sólidos domiciliarios en Perú.....	14

3.2	Influencia de la generación de residuos sólidos domiciliarios en el cómputo global de la generación de residuos municipales	17
3.3	Desperdicio de alimento (<i>food waste – FW</i>) y pérdida de alimento (<i>food loss – FL</i>) 20	
3.4	Alimentos comestibles (<i>edible food</i>) y no comestibles (<i>non-edible food</i>)	21
CAPÍTULO 4: METODOLOGÍA DE LA INVESTIGACIÓN		23
4.1	Identificación de variables influyentes en la generación de RSM	24
4.1.1.	Revisión semisistemática: Generalidades y aproximación metodológica	24
4.1.2.	Revisión semisistemática: Criterios de inclusión	25
4.1.3.	Revisión semisistemática: Estrategia de búsqueda y esquema de revisión	27
4.2	Caso de estudio.....	28
4.3	Vínculo entre la generación de RSM y el consumo de combustible.....	29
4.3.1.	Recopilación de información	29
4.3.2.	Análisis exploratorio de las variables seleccionadas	30
4.3.3.	Nexo entre la generación de residuos sólidos domiciliarios y el consumo de combustible.....	33
4.3.4.	Análisis de regresión lineal múltiple.....	33
4.3.5.	Análisis de regresión lineal múltiple: Significancia global e individual de los coeficientes del modelo	34
4.4	Implementación del modelo predictivo: <i>random forest</i>	34
4.4.1.	Recolección de datos.....	36
4.4.2.	Preprocesamiento de la información.....	38

4.4.3.	Ingeniería de variables	39
4.4.4.	Calibración de parámetros del modelo	40
4.4.5.	Indicadores de performance del modelo	40
CAPÍTULO 5: ANÁLISIS Y DISCUSIÓN DE RESULTADOS.....		41
5.1	Resultados derivados de la revisión semisistemática.....	41
5.1.1.	Consideraciones asociadas a variables socioculturales.....	41
5.1.2.	Consideraciones asociadas a variables ambientales	44
5.1.3.	Consideraciones asociadas a variables económicas.....	45
5.1.4.	Perspectiva holística del pronóstico de RSM.....	47
5.1.5.	Revisión semisistemática: Limitaciones e incertidumbres	50
5.2	Resultados derivados del análisis contextual del caso de estudio.....	51
5.3	Resultados derivados del análisis exploratorio	58
5.4	Selección de variables predictoras ajustadas al caso de investigación.....	60
5.5	Resultados derivados de la regresión lineal múltiple.....	62
5.5.1.	Escenario 1 – Combustible predominante: GLP y electricidad.....	62
5.5.2.	Escenario 2 – Consumo de gas natural preferentemente	64
5.5.3.	Escenario 3 – Combustible predominante: GLP y GN considerado como un único tipo de combustible.....	65
5.5.4.	Consideraciones adicionales	66
5.6	Resultados derivados del modelo predictivo implementado: <i>random forest</i>	67
5.6.1.	Análisis complementario de la información recopilada y balanceo de clases ...	67
5.6.2.	Optimización del modelo: Hiperparámetros	70

5.6.3.	Métricas de evaluación del modelo.....	71
5.6.4.	Importancia de las variables.....	71
5.7	Análisis de las limitaciones e incertidumbres encontradas en la ejecución metodológica	72
5.7.1.	Análisis de limitaciones	73
5.7.2.	Análisis de incertidumbres.....	74
5.8	Comparativa de los resultados obtenidos con estudios similares en la literatura.....	75
5.9	Replicabilidad de la propuesta metodológica implementada.....	77
CAPÍTULO 6: PERSPECTIVAS Y CONCLUSIONES.....		80
6.1	PI – 01: ¿Cómo podemos caracterizar a los patrones de consumo, medios y condiciones de vida de los habitantes del hogar en función a los residuos sólidos generados?	81
6.2	PI – 02: ¿Cómo podemos identificar qué tipo de variables presentan correlación con el proceso de generación de RSM?	83
6.3	PI – 03: ¿Cómo podemos medir indirectamente la generación de residuos sólidos municipales en un contexto de zonas urbanas?	84
6.4	PI – 04: ¿Cómo podemos integrar las variables recopiladas en un modelo de aprendizaje automático destinado a la predicción de residuos sólidos?.....	85
6.5	Recomendaciones.....	86
REFERENCIAS.....		87
APÉNDICE - RESÚMENES DE PRODUCCIÓN CIENTÍFICA		111
APÉNDICE – BASES DE DATOS UTILIZADAS.....		115

APÉNDICE – FIGURAS COMPLEMENTARIAS 128

APÉNDICE – EVENTOS DE DIVULGACIÓN 131



LISTA DE TABLAS

Tabla 1 - Palabras clave utilizadas durante la revisión bibliográfica sobre la influencia de las variables en el pronóstico de la generación de RSM.	26
Tabla 2 - Análisis exploratorio de las variables identificadas como impulsoras de la generación de residuos sólidos alimenticios.....	31
Tabla 3 - Variables independientes utilizadas en la predicción de residuos sólidos municipales. Fuente: Encuesta Residencial de Consumo y Usos de Energía (ERCUE) 2019-2020 (OSINERGMIN, 2021a).	37
Tabla 4 - Variables identificadas como predictoras en el proceso de generación de RSM. Adaptado de Izquierdo-Horna, Kahhat, Vázquez-Rowe, (2022).....	48
Tabla 5 - Propuesta de variables predictoras de la generación de RSM. Adaptado de Izquierdo-Horna, Kahhat, Vázquez-Rowe, (2023).....	61
Tabla 6 - Resultados del análisis de regresión múltiple para la generación per cápita de residuos alimentarios en los hogares. Adaptado de Izquierdo-Horna, Kahhat, Vázquez-Rowe, (2024)	65
Tabla 7 - Variables tratadas por presentar clases no balanceadas. Se reagruparon en nuevas categorías en función de los percentiles de cada una.....	68
Tabla A 1 - Diccionario de variables utilizadas – Sección Hogares – ERCUE 2019.....	115
Tabla A 2 - Diccionario de variables utilizadas – Sección Personas – ERCUE 2019.....	119
Tabla A 3 - Diccionario de variables utilizadas – Sección Generación de RSM - SIGERSOL 2019.....	121
Tabla A 4 - Base de datos utilizadas para relacionar el consumo de combustible doméstico y la generación per cápita de los residuos sólidos alimenticios en hogares – Variable independiente: Uso de suelo.....	122

Tabla A 5 - Base de datos utilizadas para relacionar el consumo de combustible doméstico y la generación per cápita de los residuos sólidos alimenticios en hogares – Variable independiente:

Consumo de combustibles para cocinar en el hogar..... 123

Tabla A 6 - Base de datos utilizadas para relacionar el consumo de combustible doméstico y la generación per cápita de los residuos sólidos alimenticios en hogares – Variable independiente:

Gasto en alimentos..... 125



LISTA DE FIGURAS

Figura 1 - Mapa conceptual implementado para optimizar los modelos de predicción de residuos sólidos municipales en función de una adecuada caracterización de los medios y condiciones de vida de las personas.....	6
Figura 2 - Esquematación del proceso de caracterización de residuos sólidos municipales, con énfasis en la estimación de la generación per cápita propuesto por el MINAM (2019a). 15	
Figura 3 - Representación de las fracciones comestibles y no comestibles de una piña como una forma de mostrar la distinción entre la pérdida y el desperdicio de alimentos.	22
Figura 4 - Diagrama del flujo metodológico propuesto para el pronóstico de la generación de residuos sólidos municipales.....	23
Figura 5 - Flujo metodológico de la revisión literaria realizada sobre la influencia de variables predictoras en la generación de RSM.	25
Figura 6 - Palabras clave utilizadas durante la revisión bibliográfica sobre la influencia de las variables en el pronóstico de la generación de RSM.	28
Figura 7 - Mapa de ubicación y conformación distrital del área de estudio: Lima, Perú. La coloración de los distritos corresponde a su ubicación relativa dentro de la Provincia de Lima & Provincia Constitucional del Callao (i.e., Lima Norte, Lima Sur, Lima Centro, Lima Este y Callao).....	29
Figura 8 - Representación esquemática y conceptual de una regresión lineal múltiple.	34
Figura 9 - Representación conceptual del método random forest.....	36
Figura 10 - Principales variables asociadas a la predicción de residuos sólidos.	46
Figura 11 - Variación de la incertidumbre en las estimaciones de RSM según el nivel de medición para un escenario de observación equivalente a país.	50
Figura 12 - Distribución de distritos por grupos según el porcentaje de nivel socioeconómico (NSE) en 2019, adaptado de (APEIM, 2019).	51

Figura 13 - Residuo de alimentos en hogares de Lima Metropolitana en 2019 (toneladas métricas / año). Datos tomados de SIGERSOL (MINAM, 2019b).	52
Figura 14 - Frecuencia a nivel de manzana de la fuente de energía utilizada para cocinar en la ciudad de Lima - 4a: GLP + GN, 4b: GLP, 4c: GAS NATURAL, 4d: ELECTRICIDAD. (GLP: Gas Licuado de Petróleo, GN: Gas Natural).....	54
Figura 15 - Distribución de personas que habitan en la vivienda. Información elaborada a partir de ERCUE 2019.....	55
Figura 16 - Histograma de la edad de las personas consideradas como jefe de familia. Información elaborada a partir de ERCUE 2019.....	56
Figura 17 - Histograma del gasto mensual de las personas consideradas como jefe de familia. Información elaborada a partir de ERCUE 2019.....	57
Figura 18 - Histograma referente a la generación de residuos sólidos municipales per cápita a nivel distrital de la provincia de Lima Metropolitana – 2019.....	57
Figura 19 - Análisis de regresión del consumo de combustible para cocinar en la generación doméstica de residuos alimenticios (GPC_HFW).	59
Figura 20 – Síntesis de la exploración del análisis causal de las principales variables que influyen en la generación de residuos alimentarios domésticos, basada en la evaluación realizada.	60
Figura 21 - Matriz de correlación de las variables utilizadas en el análisis de regresión múltiple: (a) Escenario 1: Consumo predominante: Gas licuado de petróleo y electricidad - (b) Escenario 2: Consumo de gas natural preferentemente - (c) Escenario 3: Consumo de combustible: GN + GLP considerado como un único tipo de combustible.	63
Figura 22 - Histograma que representa la distribución de la variable RSM_VIVIENDA agrupadas según la categoría "HM_encode". (HM_encode: Clases generadas en función del número de habitantes del hogar).....	69

Figura 23 - Histograma que representa la distribución de la generación RSM por vivienda (RSM_VIVIENDA) y el gasto mensual por persona (Expense_pers) agrupadas según la categoría "Education_level" (Education_level: Clases generadas en función del nivel de educación modal presente en la vivienda)	70
Figura 24 - Gráfico Q-Q plot que compara las distribuciones de los valores predichos y observados.....	71
Figura 25 - Importancia relativa de las variables predictoras utilizadas en el modelo de random forest para la predicción de residuos sólidos municipales.	72
Figura 26 – Propuesta de flujo de procesos para la implementación de metodologías estandarizadas para identificar variables predictivas de generación de RSM	79
Figura A 1 - Distribución espacial de los residuos sólidos domiciliarios per cápita generados a nivel distrital para Lima Metropolitana en 2019.....	128
Figura A 2 - Distribución del consumo doméstico de combustible y uso de suelo a nivel distrital para Lima Norte, Lima Este y Lima Sur.....	129
Figura A 3 - Distribución del consumo doméstico de combustible y uso de suelo a nivel distrital para Lima Centro y Callao.....	130

CAPÍTULO 1: INTRODUCCIÓN

La gestión óptima de los residuos sólidos municipales (RSM) se vuelve esencial para abordar problemas ambientales y fomentar el desarrollo socioeconómico (Zhu et al., 2021). Por ello, es importante comprender los aspectos intrínsecos de los patrones de consumo, los hábitos de producción y las dinámicas socioeconómicas presentes en la actual sociedad (Durand, 2011). Cabe resaltar que estos residuos municipales vienen generándose a un ritmo sin precedentes debido al desarrollo urbanístico y la industrialización, especialmente en países en vías de desarrollo (J. Liu et al., 2019). En ese sentido, los Objetivos de Desarrollo Sostenible (ODS) 11 y 12 de la Agenda 2030 de las Naciones Unidas emergen como pilares fundamentales en la búsqueda de soluciones integrales para la gestión de residuos sólidos y la promoción de prácticas de consumo y producción más sostenibles (ONU, 2023). Por su lado, el ODS 11, orientado hacia el desarrollo de ciudades y comunidades sostenibles, establece la necesidad imperativa de abordar la gestión de residuos sólidos como parte integral de los planes y políticas urbanas, mientras que el ODS 12 enfoca específicamente en la reducción de la generación de residuos y la promoción de un uso más eficiente de los recursos.

En este contexto, la gestión de RSM toma especial relevancia en países como el Perú, debido a las marcadas brechas existentes a nivel municipal relacionadas con la recolección, tratamiento y disposición final de los residuos sólidos (MINAM, 2019a). Por tanto, considerando que el crecimiento económico experimentado ha traído consigo un aumento considerable en la generación de RSM en la última década (Orihuela, 2018), la implementación de nuevas herramientas que faciliten la comprensión de la generación de RSM es fundamental para formular políticas públicas que se ajusten a las circunstancias y necesidades de la población.

En ese sentido, las técnicas de aprendizaje automático han surgido como una herramienta prometedora y eficaz (Abbasi & Hanandeh, 2016), debido a que la información proveniente de

los hogares exhibe cambios rápidos a corto plazo y dinámicas altamente no lineales (Cubillos, 2020). Estas circunstancias han dado lugar a un aumento significativo en su aplicación y estudio en los últimos años (Çinar et al., 2020). Por ejemplo, estos algoritmos de predicción se pueden aplicar en varios procesos de la gestión de residuos sólidos; es decir, es posible predecir la generación de RSM (Younes et al., 2015), diseñar rutas de recolección más adecuadas (Lella et al., 2017), clasificación inteligente de desechos (Sharma et al., 2019), identificación de puntos críticos (Izquierdo-Horna, Damazo, et al., 2022), entre otros.

Al emplear enfoques de análisis predictivo, se pueden anticipar las tendencias de generación de RSM en función de variables descriptivas de la sociedad y el entorno construido (Cervantes et al., 2018). Estas predicciones brindan la posibilidad de adaptar las estrategias de gestión de residuos a las necesidades cambiantes de las comunidades, contribuyendo así a un manejo más sostenible y eficiente de los RSM (Caruso & Gattone, 2019). Las técnicas más empleadas para la predicción de RSM son, principalmente, las redes neuronales artificiales (ANN), regresión lineal múltiple (MLR) y *random forest* (RF) (Assef et al., 2022).

De este modo, las redes neuronales artificiales destacan por su elevada precisión; no obstante, persiste la percepción de que estos modelos son como "cajas negras", subrayando la importancia de comprender el proceso de toma de decisiones, especialmente en áreas críticas (Burkart & Huber, 2021). Además, esta técnica requiere de más tiempo y recursos computacionales (Paulraj et al., 2016). En cuanto a la regresión lineal múltiple, este algoritmo puede prever cantidades futuras y datos más precisos de los factores más influyentes, focalizando su atención en dichos factores, pero sin abordar la predicción de todos los elementos involucrados y asumiendo una linealidad entre los factores (J. Liu et al., 2019). Por su parte, el *random forest* se presenta como un modelo adaptativo y robusto para diferentes tipos de variables, permitiendo una visión holística y precisa que no solo identifica las tendencias actuales, sino que también permite anticipar y adaptarse a los cambios dinámicos

en la generación de residuos sólidos (Iordache et al., 2022). En ese sentido, en el panorama de las técnicas de predicción, el enfoque del *random forest* destaca por su versatilidad y eficacia, abordando de manera integral la complejidad de los datos relacionados con la generación de residuos sólidos (Burkart & Huber, 2021).

En consecuencia, considerando que este proyecto de investigación tiene como objetivo principal pronosticar la generación de RSM haciendo uso de variables sociales, económicas y ambientales, se utilizó un modelo de *random forest* para determinar y evaluar el potencial predictivo de las variables seleccionadas. Para lograrlo, en una primera etapa se llevó a cabo una revisión semisistemática para analizar la influencia de cada una de estas variables en la generación de RSM. En una segunda fase, se eligió Lima Metropolitana (que incluye la Provincia Constitucional del Callao) como caso de estudio para identificar las variables más influyentes en la predicción de RSM domiciliarios. Finalmente, se realizó la predicción a nivel vivienda en función de los datos recopilados.

CAPÍTULO 2: PLANTEAMIENTO DE LA INVESTIGACIÓN

2.1 Diagnóstico del contexto actual

El crecimiento exponencial de la generación de RSM representa uno de los principales desafíos a nivel global (Espinoza Pérez et al., 2021). Para abordar este problema, es necesario tener una visión enfocada en aprovechar los recursos de manera eficiente, lo cual implica desarrollar estrategias que prioricen la reducción, reutilización y reciclaje de los RSM para minimizar su impacto ambiental (Kaur et al., 2021). Al adoptar estas medidas, se promueve una gestión más eficiente y sostenible de los recursos, lo que impulsa el avance hacia una economía circular (Agovino et al., 2024). Es fundamental implementar acciones que maximicen el aprovechamiento de los recursos contenidos en los residuos, contrarrestando así los efectos negativos del aumento en su producción (Blasenbauer et al., 2024). Este escenario se ve condicionado también por factores como políticas públicas poco efectivas, falta de conciencia ambiental y el rápido crecimiento poblacional (EPA, 2020). Esto último sugiere que, a medida que las ciudades se desarrollan, la cantidad de residuos sólidos urbanos crece incluso de manera más veloz que la tasa de urbanización (Jordán et al., 2017). Por lo tanto, es necesario una evaluación holística que contribuya a mejorar las políticas existentes para hacer frente a este problema de manera efectiva (Dou & Toth, 2021).

Además, considerando la significativa contribución de los hogares a la cantidad total de RSM, es plausible afirmar que una parte sustancial de estos residuos puede estar relacionada directamente con la gestión de la pérdida y desperdicio de materia orgánica, especialmente de alimentos (González-Santana et al., 2022). En tal sentido, se comprende que el desperdicio de alimentos es un problema multifacético que requiere la integración de diferentes perspectivas disciplinarias (Schanes et al., 2018). Sin embargo, la disponibilidad de información de calidad sobre este aspecto es limitada y recientes investigaciones han identificado inconsistencias metodológicas en su cuantificación (Hoehn et al., 2022), lo que se considera uno de los

principales obstáculos para abordar dicha problemática. Por ende, resulta imperativo establecer metodologías estandarizadas para su cuantificación en hogares (Withanage et al., 2021).

En este contexto, es importante reconocer que los hogares desempeñan un papel clave en la formulación de estrategias de mitigación y adaptación, especialmente en lo que respecta al consumo residencial de energía /combustible (Cárdenas-Mamani et al., 2022). La información sobre el consumo residencial de combustible puede proporcionar una alternativa valiosa para comprender mejor los patrones de consumo y su relación con la generación de RSM. Por ejemplo, un análisis del consumo de combustible para cocinar en hogares residenciales en Lima, Perú, señaló que el gas licuado de petróleo (GLP) es la principal fuente de energía en los hogares de bajos ingresos (OSINERGMIN, 2021b). Además, se observó un crecimiento significativo en el consumo de gas natural (GN), que se ha vuelto más accesible para el sector residencial debido a su menor costo en comparación con el GLP (OSINERGMIN, 2020a). En contraste, el uso de electricidad para cocinar es menos frecuente, principalmente debido al alto costo asociado con su consumo en comparación con otras fuentes de combustible disponibles.

2.2 Abordando la complejidad en el pronóstico de RSM

El dinamismo constante en los patrones de consumo y en las características de los ciudadanos, sumado al acelerado crecimiento poblacional y expansión urbana desordenada, genera un escenario de gestión complejo en el ámbito de los RSM. Esta complejidad requiere de una innovación conjunta no solo a nivel conceptual, sino también en términos prácticos y metodológicos. En ese sentido, tras identificar previamente las falencias asociadas al pronóstico de los RSM —especialmente en el proceso de recolección de información, procesamiento y tratamiento de datos—, resulta fundamental proponer estrategias replicables que sirvan como línea base para una adecuada caracterización del caso de estudio. Esta propuesta conceptual se resume en la Figura 1, la cual subraya la importancia de comprender las dinámicas internas de una comunidad para abordar la generación de RSM. Para lograr este

entendimiento, se sugiere la aplicación de técnicas como el análisis exploratorio y revisiones literarias. Asimismo, se evidencia la necesidad de contar con un conjunto eficiente de variables que posean una capacidad adecuada de segmentación, las cuales deben provenir de fuentes de información robustas y confiables.



Figura 1 - Mapa conceptual implementado para optimizar los modelos de predicción de residuos sólidos municipales en función de una adecuada caracterización de los medios y condiciones de vida de las personas.

2.3 Definición del problema de investigación

Dado el crecimiento significativo en la producción de RSM en los últimos años, las proyecciones actuales indican que, para el año 2025, la generación anual de RSM alcanzará la cifra de 1.42 kg per cápita por día (2.2 mil millones de toneladas métricas) (Hoornweg & Bhada-Tata, 2012). Además, se prevé que, para el año 2050, esta cifra aumentará considerablemente llegando a 3.4 mil millones de toneladas métricas, ejerciendo una presión abrumadora sobre los sistemas de gestión de residuos existentes (Kaza et al., 2018). Por ello,

predecir la cantidad de RSM generados en un determinado período de tiempo y alcance geográfico se vuelve especialmente relevante en la gestión de RSM (Dyson & Chang, 2005).

Como resultado, se han desarrollado diferentes estrategias de análisis, que van desde enfoques tradicionales como el análisis de regresión (Kumar & Samadder, 2017), el cual consiste en hacer suposiciones sobre la relación entre dos o más variables explicativas y una variable de respuesta (Azadi & Karimi-Jashni, 2016), hasta enfoques de aprendizaje automático más complejos que pueden identificar patrones en los datos observados y hacer predicciones sin ser programados explícitamente (Nguyen et al., 2021). Esto último permite la implementación de modelos como redes neuronales (Abbasi & Hanandeh, 2016), árboles de decisión (Kannangara et al., 2018), máquinas de soporte vectorial (Abdallah, Talib, et al., 2020), entre otros.

Las predicciones de generación de RSM son posibles a través de extrapolaciones basadas en representaciones reconocidas de la situación actual, validándolas a través de la robustez de las variables elegidas para este propósito (Giampietro & Mayumi, 2018). Estas conceptualizaciones son factibles de establecer porque la generación de RSM está directamente relacionada con las actividades diarias de una ciudad específica (Vyas et al., 2022). Este proceso predictivo se puede explicar a través de las relaciones entre las variables socioculturales, ambientales y económicas que caracterizan la realidad (Sunayana et al., 2021). La identificación oportuna de estas variables que influyen en la generación de RSM permitirá generar un escenario para una gestión eficiente de residuos (Ordoñez, 2004) y su aprovechamiento (Scarlat et al., 2015).

En ese sentido, estos estudios intentan determinar un conjunto adecuado de variables capaces de predecir la generación de RSM en función del comportamiento y las condiciones de vida de las personas en un contexto urbano específico. Sin embargo, esta tarea es desafiante debido a su variabilidad en el espacio y el tiempo (Chhay et al., 2018). Además, la disponibilidad y

confiabilidad de los datos desempeñan un papel crítico, no solo en la determinación de estas variables clave, sino también en el establecimiento de los canales de acceso a la información (B. Liu et al., 2021). Algunos estudios utilizan censos (Kollikkathara et al., 2010), bases de datos proporcionadas por organizaciones relacionadas (Serrona & soo Yu, 2009), encuestas o datos geográficos (Paulauskaite-Taraseviciene et al., 2022) para recopilar datos primarios y superar esta brecha.

Por otro lado, esta investigación ha identificado dos vacíos importantes a partir de la revisión de la literatura existente sobre el tema. El primero es la falta de preprocesamiento para determinar e identificar variables que caractericen a la población y el segundo es el grado de disponibilidad de los datos para la implementación de un modelo estadístico posterior. Aunque existen estudios previos sobre el análisis de ciertas variables en el proceso de generación de RSM (Hannan et al., 2020), hasta la fecha no se ha evaluado de manera integral la influencia de variables socioculturales, ambientales y económicas en la generación de RSM con el objetivo de optimizar los modelos predictivos.

Finalmente, con la finalidad de minimizar las brechas identificadas y contribuir a una eficiente gestión de RSM, nos planteamos la siguiente pregunta de investigación:

¿Cómo podemos predecir los RSM cuando la disponibilidad de información es limitada y los recursos económicos son escasos?

Esta pregunta de investigación se extiende a cuatro preguntas de investigación específicas que ayudarán a comprender los objetivos de este proyecto.

PI 1 – ¿Cómo podemos caracterizar a los patrones de consumo, medios y condiciones de vida de los habitantes del hogar en función a los residuos sólidos generados?

PI 2 – ¿Cómo podemos identificar qué tipo de variables presentan correlación con el proceso de generación de RSM?

PI 3 – ¿Cómo podemos medir indirectamente la generación de RSM en un contexto de zonas urbanas?

PI 4 – ¿Cómo podemos integrar las variables recopiladas en un modelo de aprendizaje automático destinado a la predicción de residuos sólidos?

En ese sentido, con la finalidad de responder las preguntas anteriormente planteadas, se han formulado las siguientes hipótesis de investigación requeridas para el desarrollo conceptual y metodológico:

Hipótesis 1 – Las dimensiones sociales, económicas y ambientales ofrecen robustez y significancia para caracterizar los patrones de consumo, medios y condiciones de vida de los residentes.

Hipótesis 2 – Un análisis de regresión lineal múltiple permitirá identificar si existe o no significancia al relacionar la variable elegida y la cantidad de residuo generado.

Hipótesis 3 – El uso de combustible para cocinar se presenta como un indicador robusto y fiable para medir indirectamente el residuo generado.

Hipótesis 4 – Dada la naturaleza de las variables, un modelo de *random forest*, se ajusta mejor a los objetivos de este estudio.

2.4 Objetivos de la investigación

El objetivo principal de esta investigación es implementar un modelo de predicción de RSM a nivel vivienda utilizando información de acceso libre, especialmente aquella relacionada a las características propias de cada hogar.

Para tal fin, se plantearon los siguientes objetivos específicos:

- Analizar los patrones de consumo en los hogares urbanos en función de la literatura existente y de los datos disponibles.

- Identificar un conjunto de variables con el potencial de predecir la generación de RSM mediante la caracterización de las condiciones de vida y el entorno construido.
- Examinar la correlación existente entre la generación de RSM y el uso de combustibles para cocinar.
- Diseñar y modelar un modelo de aprendizaje automático para la predicción de la generación de RSM a nivel distrital utilizando datos de acceso público (i.e., censos y encuestas gubernamentales).

2.5 Alcance, justificación y limitaciones

2.5.1. Alcance de la investigación

Este proyecto se enfoca exclusivamente en el análisis de los aspectos vinculados a la generación de RSM en entornos urbanos. Concretamente, se concentra en la evaluación del impacto de variables específicas en la producción de residuos. Se utilizarán, como referencia, los datos conseguidos del año 2019, ello considerando las complicaciones y alteraciones causadas por la aparición de la pandemia por la COVID-19. La principal fuente de información provendrá de bases de datos gubernamentales de acceso público, particularmente aquellas relacionadas con el consumo de combustibles en entornos residenciales. En cuanto a la dimensión geoespacial, el estudio se circunscribe a Lima Metropolitana (que incluye la Provincia Constitucional del Callao). Finalmente, vale recalcar que la metodología empleada se presenta como un modelo replicable, sujeto a la validación de las condiciones específicas de cada caso de estudio.

2.5.2. Justificación de la investigación

Esta investigación se justifica en términos prácticos al enfrentar una problemática urgente vinculada a la gestión de residuos sólidos en áreas urbanas del ‘Sur Global’¹. Se enfoca de

¹ Sur global: Hace referencia generalmente a los países menos desarrollados o en vías de desarrollo, ubicados principalmente en el hemisferio sur del planeta (Valente de Macedo et al., 2021).

manera específica en el pronóstico de estos residuos a través de la utilización de técnicas de aprendizaje automático. La eficaz gestión de estos residuos resulta fundamental tanto para evitar consecuencias ambientales adversas como para optimizar la utilización de recursos valiosos. Por otro lado, en términos teóricos, este proyecto aporta a la comunidad científica nuevas herramientas para la recolección y procesamiento de datos relacionados a variables potencialmente capaces de explicar la generación de RSM. Además, propone un conjunto de variables que pueden ser utilizadas para tal fin u otros procesos afines a la gestión de RSM. Finalmente, en términos metodológicos, este proyecto tiene como objetivo validar el uso de técnicas de aprendizaje automático para predecir la generación de RSM, haciendo uso de variables cuidadosamente seleccionadas. Se propone la aplicación de un enfoque que combine el análisis de datos históricos y demográficos con la utilización de algoritmos de predicción.

La selección de variables para la implementación del modelo predictivo incluye el análisis y entendimiento de patrones históricos de generación de residuos, características demográficas y urbanas relevantes, entre otros factores pertinentes. La inclusión de estas variables tiene como objetivo mejorar la capacidad predictiva del modelo, proporcionando no solo estimaciones cuantitativas, sino también indicios significativos sobre los determinantes subyacentes de la generación de RSM.

2.5.3. Limitaciones de la investigación

Dentro de las limitaciones identificadas en la ejecución de este proyecto, resalta principalmente la ausencia de datos actualizados y confiables sobre la generación de residuos sólidos domiciliarios. Esta carencia se agrava aún más debido a la escasa divulgación y actualización en los procesos de cuantificación y caracterización de estos residuos por parte de las organizaciones gubernamentales responsables de dicho proceso (p. ej., MINAM). Otra limitación relevante está relacionada con la densidad poblacional en algunos de los distritos analizados; es decir, el número de muestras disponibles es mínimo, lo que compromete la

representatividad de la información recopilada. La escasez de datos en estas áreas específicas podría distorsionar los resultados y limitar la generalización de las conclusiones obtenidas, lo cual constituye una consideración importante al interpretar los hallazgos de la investigación.

En términos de alcance, es fundamental destacar que esta investigación no aborda el impacto generado por servicios específicos, como el servicio de entrega de alimentos, ni considera suelos residenciales con usos distintos a la vivienda. Esta omisión puede limitar la comprensión integral de la generación de residuos sólidos en entornos urbanos, ya que ciertos aspectos, como los patrones de consumo asociados al servicio de traslado de comida, podrían tener un impacto significativo en la cantidad y composición de los residuos generados.

2.6 Estructura de la tesis y producción científica

Esta sección desarrolla el marco conceptual que rige este estudio. Así pues, sentará las bases necesarias para diferenciar algunos conceptos importantes utilizados en esta investigación. En la subsiguiente sección responderemos a las PI-1 y PI-2 a través de la revisión semisistemática realizada. Asimismo, probamos la Hipótesis 1 a través de cotejo del diagnóstico exploratorio de la zona de estudio y la literatura disponible presentada en el capítulo 1 y 2. Los contenidos mostrados en esta sección (tablas y figuras) son —en su gran mayoría— nuestros resultados publicados en Izquierdo-Horna et. al., (2022). En el subsiguiente capítulo, son abordadas las PI-3 y PI-4, las cuales presentan como potencial indicador de generación de RSM al consumo de combustible para cocinar. Además, se describen las características clave del caso de estudio y se exponen los principales descubrimientos resultantes de la aplicación de esta metodología. Posteriormente, se procede a analizar los resultados más relevantes de este proyecto de tesis y a presentar las conclusiones obtenidas.

Por otro lado, en el marco de este proyecto, se ha generado una producción científica que incluye los siguientes trabajos. En primer lugar, se encuentra el artículo titulado "Revisión de

la influencia de variables socioculturales, ambientales y económicas para predecir la generación de residuos sólidos municipales (RSM)", el cual ha sido publicado en la revista "*Sustainable Production and Consumption*" (<https://doi.org/10.1016/j.spc.2022.08.008>). Este documento contempla la primera parte de este proyecto de investigación y se enfoca en explorar las distintas variables asociadas a un modelo predictivo de RSM. Este documento responde a las PI-1 y PI-2.

Además, se presentó la ponencia en la conferencia *SARDINIA 2023* con el título "Identificación de variables relacionadas con la predicción de la generación de residuos sólidos domésticos en áreas residenciales" (<https://www.sardiniasymposium.it>). De la misma forma, este trabajo representa el segundo componente de este estudio y su alcance compete a la identificación del conjunto de variables y dimensiones recomendables para un modelo predictivo, contemplando las características del caso de estudio. Este documento responde a la PI-3 y sienta las bases teóricas para responder a la PI-4.

Como tercer hito, se encuentra el artículo titulado "Descubriendo la relación entre el consumo de energía y el desperdicio de alimentos en los hogares: Un enfoque en los principales predictores de la generación de desperdicios de alimentos", el cual ha sido publicado en la revista "*Journal of Material Cycles and Waste Management*" (<https://doi.org/10.1007/s10163-024-01946-2>). El alcance de este manuscrito compete la validación de las variables escogidas como predictores de RSM. Finalmente, se utilizaron los resultados preliminares de una investigación —aún por publicar— con los resultados de la implementación del modelo *random forest*. Este documento responde a la PI-4 y valida las Hipótesis 3 e Hipótesis 4.

CAPÍTULO 3: MARCO CONCEPTUAL DE LA INVESTIGACIÓN

El concepto de RSM ha sido objeto de extenso estudio por diversas organizaciones y varios investigadores han propuesto sus propias definiciones. Por ejemplo, según la Organización para la Cooperación y el Desarrollo Económicos (OCDE), los RSM son los residuos recolectados por o en nombre de municipios e incluyen residuos domésticos y similares generados por diversas actividades que disponen de sus residuos en las mismas instalaciones (p. ej., vertederos a cielo abierto o rellenos sanitarios) utilizadas para la recolección de residuos por parte de los municipios (OCDE, 2015). Para las Naciones Unidas, una definición funcional de los RSM apunta a "*residuo generado por hogares y de naturaleza similar, con la excepción de residuos peligrosos y residuos de procesos industriales*" (UN-HABITAT, 2010). En la misma línea, el Banco Mundial (Kaza et al., 2018) define los RSM como los residuos provenientes de fuentes residenciales, comerciales e institucionales. Finalmente, Lagerkvist y Dahlén (2012) complementan estas definiciones al enfatizar que el término "residuo" se comprende como la consecuencia inevitable de acciones donde se utilizan materiales y no se gestionan adecuadamente.

Bajo estos conceptos, resulta necesario comprender las distintas relaciones intrínsecas entre las comunidades humanas y el entorno construido traducidas en variables predictoras. Así pues, la predicción de la generación de RSM es un proceso complejo debido a la alta y persistente incertidumbre asociada y a la influencia que los resultados tienen en la toma de decisiones (Sunayana et al., 2021). Por lo tanto, el proceso de selección de variables predictoras es crítico y debe ir acompañado de juicio experto debido a la incompatibilidad de ciertas dimensiones y la pérdida de información asociada (Munda, 2014).

3.1 Caracterización y valorización de los residuos sólidos domiciliarios en Perú

La caracterización de RSM representa un insumo fundamental en la elaboración de instrumentos para su propia gestión, formulación de proyectos de inversión; entre otros (Chang

& Davila, 2008). En ese sentido, el Ministerio del Ambiente diseñó una guía estandarizada para caracterizar a los RSM mediante una metodología que describe de manera práctica los pasos para la obtención de cifras locales relacionadas a este tipo de residuos (MINAM, 2019a). Así pues, implementar eficazmente un estudio de caracterización de RSM implica tres etapas fundamentales: planificación, trabajo de campo y análisis de información. En la primera etapa, se lleva a cabo la organización del distrito en zonas de estudio y se distribuye la muestra de manera equitativa y aleatoria. La segunda etapa implica el trabajo de campo, donde se realizan los procedimientos para la participación de los predios del estudio, manejo, análisis y traslado de muestras. Finalmente, en la tercera etapa, se lleva a cabo el análisis de la información, obteniendo los índices deseados (MINAM, 2019a). La Figura 2 ilustra el proceso sugerido para la caracterización de RSM.



Figura 2 - Esquemática del proceso de caracterización de residuos sólidos municipales, con énfasis en la estimación de la generación per cápita propuesto por el MINAM (2019a).

De la misma forma, la guía también establece que los estudios de caracterización de RSM deben realizarse cada 5 años, debiendo evitarse períodos de cambios sustanciales en la generación de residuos, ya sea un aumento o una disminución significativa. Esto se hace con el fin de garantizar la fiabilidad de los datos recopilados y la efectividad de las estrategias de gestión de residuos. En caso de que sea necesario realizar un estudio de caracterización durante

tales períodos de variación significativa, se deberá llevar a cabo más de una evaluación. Asimismo, las autoridades municipales deben tener en cuenta consideraciones relevantes como el crecimiento demográfico, la aplicación de estrategias para la minimización de residuos sólidos y desarrollo o aumento de actividades económicas dentro de sus respectivos distritos (MINAM, 2019a).

Por otro lado, a nivel internacional, la estimación y caracterización de RSM se lleva a cabo mediante diversos métodos y técnicas estandarizadas (Magrinho et al., 2006). En Viena Austria, por ejemplo, se establecen sistemas de colección basados en puntos de recogida y densidad poblacional, siendo el método ‘puerta-puerta’ el más común (EEA, 2022). Esto se realiza con la finalidad de evaluar el potencial de los recursos (Kladnik et al., 2024). De la misma forma, en California, el Departamento de Recursos de Reciclaje y Recuperación (*CalRecycle*) lidera un proceso de caracterización de residuos que implica la recolección representativa de muestras de hogares, empresas y estaciones de transferencia. Estos datos son luego analizados para identificar tendencias y patrones en la composición de los residuos, informando así la formulación de políticas y programas de gestión de residuos efectivos y sostenibles (CalRecycle, 2024).

En cambio, en países emergentes o en vías de desarrollo, el proceso de caracterización de RSM está poco documentado y a menudo no se sigue un lineamiento común. Por ejemplo, en la India, Patwa et al., (2020) concuerdan en que es fundamental recabar una mayor cantidad de información sobre la producción y composición de residuos sólidos, debido a las variaciones en la geografía, el clima, la cultura, los patrones alimentarios y los métodos actuales de gestión de residuos sólidos. En esta misma línea, Abdallah et al., (2020), concuerdan que en Egipto también hacen falta datos confiables y actualizados sobre las características de los residuos sólidos municipales, así como sobre su correlación con diferentes parámetros locales generando un gran impacto en la economía de los procesos de gestión de residuos.

Similarmente, en Uganda, los esfuerzos para gestionar los RSM en la ciudad se ven constantemente abrumados por el crecimiento continuo de la población urbana, los niveles incrementados de actividad económica y la reducción de fondos por parte del gobierno central (Komakech et al., 2014)

En términos generales, una correcta caracterización de RSM desempeña un papel fundamental en la gestión adecuada de estos residuos, al proporcionar información relevante que permite optimizar la recuperación de recursos y valorizar los materiales desechados en todo el espectro de residuos (Hemali-Arosha et al., 2024). Esto, a su vez, facilita la generación de productos de valor a partir de los residuos y la creación de oportunidades económicas mediante la valorización de los RSM (Kaur et al., 2021). Además, según Li & Skelly (2023), la gestión de RSM puede evaluarse de diversas formas, como la transformación de los residuos en recursos valiosos a través de la reutilización, el reciclaje y la valorización energética, contribuyendo así a la economía circular y reduciendo la dependencia de los vertederos. Por otro lado, una caracterización imprecisa de los RSM puede provocar daños al ambiente y a la salud pública en caso de acumulación, lo que conduce a una gestión inadecuada de los mismos (Ibarra-Esparza et al., 2023).

3.2 Influencia de la generación de residuos sólidos domiciliarios en el cómputo global de la generación de residuos municipales

El crecimiento descontrolado de los RSM también está estrechamente vinculado a la falta de recursos e infraestructuras apropiadas para su gestión (Powell et al., 2018). Esta situación genera una serie de desafíos adicionales, entre ellos la falta de información completa y transparente; así como de una gobernanza limitada (Cervantes et al., 2018). Además, debido a las condiciones demográficas y económicas existentes, la composición de los RSM influye en las estrategias utilizadas para su gestión (Karak et al., 2012). En este contexto, investigaciones previas han identificado que los residuos sólidos domésticos representan una fracción

significativa del total de RSM (Suthar & Singh, 2015), y que su composición está influenciada principalmente por el consumo de bienes domésticos (Abdel-Shafy & Mansour, 2018), demanda de energía (Tomić & Schneider, 2017) o el uso de recursos disponibles, así como aspectos políticos (p. ej., políticas y regulaciones gubernamentales), técnicos (p. ej., infraestructura y tecnologías de gestión de residuos) y legales (p. ej., leyes y regulaciones de disposición de residuos) (Al-Khatib et al., 2010).

Dado que los hogares desempeñan un papel clave en la formulación de estrategias de mitigación y adaptación al impacto ambiental en entornos urbanos (Cárdenas-Mamani et al., 2022), resulta fundamental comprender cómo los patrones de consumo domiciliario impactan en la generación de residuos sólidos para mejorar las políticas de gestión de RSM (Deus et al., 2020). Para lograr esto, la implementación de estrategias que fomenten el desarrollo competitivo (Aja & Al-Kayiem, 2014) y sostenible (Vázquez-Rowe et al., 2019) es fundamental. Al alinear estos enfoques, se puede establecer un marco integral para optimizar las prácticas de gestión de RSM y minimizar los impactos ambientales. Sin embargo, en muchos casos, esto aún no se ha implementado con éxito debido a un enfoque persistente y sistémico, lo que resulta en una gestión ambiental débil (Orias et al., 2021).

Cuando se analiza con mayor detalle la composición de los residuos sólidos domésticos, estos están compuestos por una amplia variedad de tipos de residuos de diferente naturaleza química y biológica, siendo la fracción orgánica la más predominante (Sujauddin et al., 2008), especialmente en ciudades emergentes y en desarrollo (Azevedo et al., 2021). Aunque en los hogares se pueden encontrar otras fuentes de residuos orgánicos (p. ej., residuos de jardín), los residuos de alimentos están directamente relacionados con la fracción orgánica (González-Santana et al., 2022). Así, los desperdicios de alimentos, producidos desde la fase de distribución hasta el consumo, constituyen un problema complejo y multifacético que requiere la integración de diferentes perspectivas multidisciplinarias (Schanes et al., 2018) y representa

un desafío ambiental global urgente según los Objetivos de Desarrollo Sostenible (ODS), como el ODS 11, que se refiere a ciudades y comunidades sostenibles, y el ODS 12, que se relaciona con el consumo y la producción responsables (Naciones Unidas, 2021). Sin embargo, los desperdicios de alimentos también encierran un potencial significativo para ser aprovechados de manera efectiva. Por ejemplo, el compostaje es una estrategia destacada para reutilizar los desperdicios de alimentos en la agricultura (Morone et al., 2019); también, se pueden utilizar como alimento para ganado o como materia prima para la producción de biogás a través de la digestión anaeróbica (Rodríguez-Jiménez et al., 2022).

El impacto del sector alimentario en los problemas de gestión de residuos de alimentos en los hogares representa un desafío importante (Vázquez-Rowe et al., 2021), ya que varía con el tiempo como resultado de los desarrollos socioeconómicos variables (Springmann et al., 2018). Esta variabilidad se vuelve significativa en términos de aumento de la generación de residuos de alimentos en los hogares, vinculada a cambios en el consumo de energía (Laso et al., 2018) y características demográficas (i.e., uso del suelo y aspectos socioeconómicos) (Ma & Hipel, 2016). Por lo tanto, para lograr estrategias efectivas para gestionar los residuos alimentarios del hogar, es esencial establecer e implementar una base de datos adecuada (Chhay et al., 2018) y estandarizar metodologías de cuantificación para la comparabilidad (Hoehn et al., 2023). Los hallazgos de un estudio sobre la generación de residuos domésticos en Medan (Indonesia) (Khair et al., 2019) enriquecen aún más nuestra comprensión de estos desafíos, enfatizando la necesidad de estrategias de datos sólidas para abordar las complejidades de una gestión efectiva de los residuos alimentarios, resaltando los desafíos de precisión de datos y consistencia metodológica.

De esta manera, para mejorar las estrategias de gestión de residuos en países en desarrollo, es imperativo realizar una revisión detallada de los factores que influyen en la generación de residuos alimentarios del hogar y de RSM (Dou & Toth, 2021). Las percepciones obtenidas de

estudios en Nablus (Palestina) (Pirani et al., 2015) y Bangkok (Tailandia) (Sukholthaman et al., 2017) destacan las conexiones intrincadas entre elementos socioeconómicos y demográficos, arrojando luz sobre la dinámica de la generación de residuos a nivel domiciliario. En conjunto, estos estudios enriquecen la literatura científica sobre el tratamiento de los residuos y aportan nuevas perspectivas que sirven como referentes relevantes para la elaboración de estrategias personalizadas en la gestión de los mismos.

3.3 Desperdicio de alimento (*food waste – FW*) y pérdida de alimento (*food loss – FL*)

Durante los últimos años, el concepto de pérdida y desperdicio de alimentos (*food loss and waste - FLW*) ha venido ganando importancia debido a su impacto a lo largo de toda la cadena alimentaria, desde la fase inicial de producción agrícola hasta el momento del consumo final en los hogares (FAO, 2014). Esta preocupante problemática se refleja en las estadísticas globales, ya que alrededor de un tercio de la producción mundial de alimentos destinados al consumo humano se pierde o desperdicia anualmente, lo que se traduce en aproximadamente 1 300 millones de toneladas métricas (FAO, 2011). Esta cifra representa aproximadamente el 38 % de la energía total necesaria para llevar a cabo la producción de dichos alimentos (Hoehn et al., 2022).

En países con un nivel socioeconómico medio-alto, se evidencia un significativo desperdicio de alimentos durante la fase de consumo, implicando el descarte de productos aún aptos para el consumo humano. En contraste, en países de bajos ingresos, las pérdidas de alimentos se concentran principalmente en las etapas iniciales e intermedias de la cadena de suministro alimentaria, con una menor incidencia de desperdicio en la etapa de consumo (FAO, 2011). Por lo tanto, para un mejor proceso de cuantificación de pérdidas y desperdicios de alimentos, resulta prudente definir y diferenciar ambos conceptos e identificar las fases que involucra. Por un lado, la pérdida de alimentos se caracteriza por la reducción de la cantidad o calidad de la masa de alimentos aptos para el consumo humano a lo largo de la cadena de suministro que

conduce específicamente a dichos alimentos (Hoehn et al., 2023). Estas pérdidas alimenticias se producen en las etapas de producción, poscosecha y procesamiento dentro de la cadena de suministro de alimentos (FAO, 2011). También incluye todos los usos no alimentarios, como los alimentos desechados que originalmente estaban destinados al consumo humano y luego se reciclaron como alimentos para animales (García-Herrero et al., 2018).

Por otro lado, el desperdicio de alimentos constituye una parte única y separada de la pérdida de alimentos debido a que tiene un tratamiento diferente (Isah & Ozbay, 2020). Este término hace referencia a la eliminación de alimentos que aún son aptos para el consumo, ya sea por decisión propia o porque se han dejado caducar por negligencia, en la mayoría de los casos, del consumidor final (FAO, 2014). Además, este tipo de desperdicio está vinculado al comportamiento de los consumidores (van Dooren et al., 2020). En líneas generales, el desperdicio de alimentos se entiende como la fracción desechada de alimentos que se genera al final de la cadena de suministro de alimentos (Hoehn et al., 2023).

3.4 Alimentos comestibles (*edible food*) y no comestibles (*non-edible food*)

Para comprender con más detalle el concepto de pérdida y desperdicio de alimento, introduciremos dos conceptos adicionales para un mejor entendimiento. En ese sentido, la FAO (2011) distingue entre alimentos comestibles (*edible food*) y no comestibles (*non-edible food*); sin embargo, es importante señalar que ciertas fracciones pueden considerarse comestibles o no comestibles por diversas razones, como diferencias culturales o posibles riesgos para la salud (Mladenova & Shtereva, 2009). De este modo, el término ‘no comestible’ hace referencia a las partes del alimento que socialmente se consideran no aptas para el consumo humano (Hoehn et al., 2023).

En ese sentido, se puede decir que la pérdida de alimento (FL) se puede cuantificar en función a la disminución de la masa de alimentos comestibles como no comestibles a lo largo de la

cadena de suministro; mientras que el desperdicio de alimento (FW), en función de la fracción comestible. Finalmente, la Figura 3 muestra la representación visual de las fracciones comestibles y no comestible asociadas al desperdicio o pérdida de alimentos.



Figura 3 - Representación de las fracciones comestibles y no comestibles de una piña como una forma de mostrar la distinción entre la pérdida y el desperdicio de alimentos.

La flecha hacia arriba indica un aumento en la condición indicada; mientras que, la flecha hacia abajo, indica una disminución.

CAPÍTULO 4: METODOLOGÍA DE LA INVESTIGACIÓN

La propuesta metodológica empleada, en este trabajo de investigación, se muestra en la Figura 4 y se desarrolla de manera secuencial en esta sección. La metodología involucra una revisión semisistemática de los últimos estudios asociados al pronóstico de la generación de RSM comprendida desde el año 2000 hasta el 31 de julio de 2022. Posteriormente, a través de juicio de expertos, se validan e identifican las potenciales variables predictoras contemplando la disponibilidad de datos y confiabilidad de la misma. Finalmente, se evalúa la técnica de aprendizaje automático que más se ajuste a los datos disponibles.



Figura 4 - Diagrama del flujo metodológico propuesto para el pronóstico de la generación de residuos sólidos municipales.

4.1 Identificación de variables influyentes en la generación de RSM

4.1.1. Revisión semisistemática: Generalidades y aproximación metodológica

Con la finalidad de explorar las distintas combinaciones de variables que pueden influir en el proceso de generación de RSM, esta revisión semisistemática tiene como objetivo examinar la influencia de las variables predictoras en el proceso de generación de RSM, así como explicar si se eligen por su capacidad predictiva, disponibilidad de datos o una combinación de ambas. Además de proporcionar una descripción completa de cada dimensión, considerando las posibles interrelaciones entre ellas. Los documentos considerados para esta revisión son aquellos publicados desde el año 2000 hasta el 31 de julio de 2022 sobre predicciones de generación de residuos sólidos basadas en variables socioculturales, ambientales y económicas seleccionadas mediante criterios de inclusión y exclusión. Esta sección muestra el marco analítico de la revisión realizada siguiendo las recomendaciones generales para la elaboración de una revisión crítica de la literatura (Xiao & Watson, 2019).

La literatura subraya que las características y composición de los RSM varían no solo según la ubicación, sino también según factores socioeconómicos, climáticos y de mercado (Chand Malav et al., 2020). Además, algunos estudios enfatizan también la influencia de las tecnologías de tratamiento de RSM, la recuperación de energía y las consideraciones en la cadena de suministro, entre otros aspectos técnicos, sociales, económicos y ambientales (Gaviria-Cuevas et al., 2019). A pesar de que existen estudios previos sobre el análisis de ciertas variables en el proceso de generación de RSM (Hannan et al., 2020), hasta la fecha no se ha evaluado de manera integral la influencia de variables socioculturales, ambientales y económicas para lograr una optimización de los modelos predictivos.

La Figura 5 presenta el proceso metodológico seguido, incluido el flujo de trabajo llevado a cabo para desarrollar esta investigación. El primer paso consistió en la definición de la pregunta de investigación y, posteriormente, en la determinación de los criterios de inclusión y

exclusión. Estos criterios fueron establecidos por los autores en función de la línea temporal analizada, las bases de datos y la relevancia para el tema. Un total de 178 artículos cumplían con los criterios establecidos. Posteriormente, los artículos fueron sometidos a un proceso de análisis individual basado en sus características específicas (p. ej., título, resumen, palabras clave, etc.) para determinar su relevancia para el tema de interés. La siguiente etapa implicó una lectura en profundidad de 47 artículos, de los cuales se seleccionaron 36 para su inclusión en la revisión.

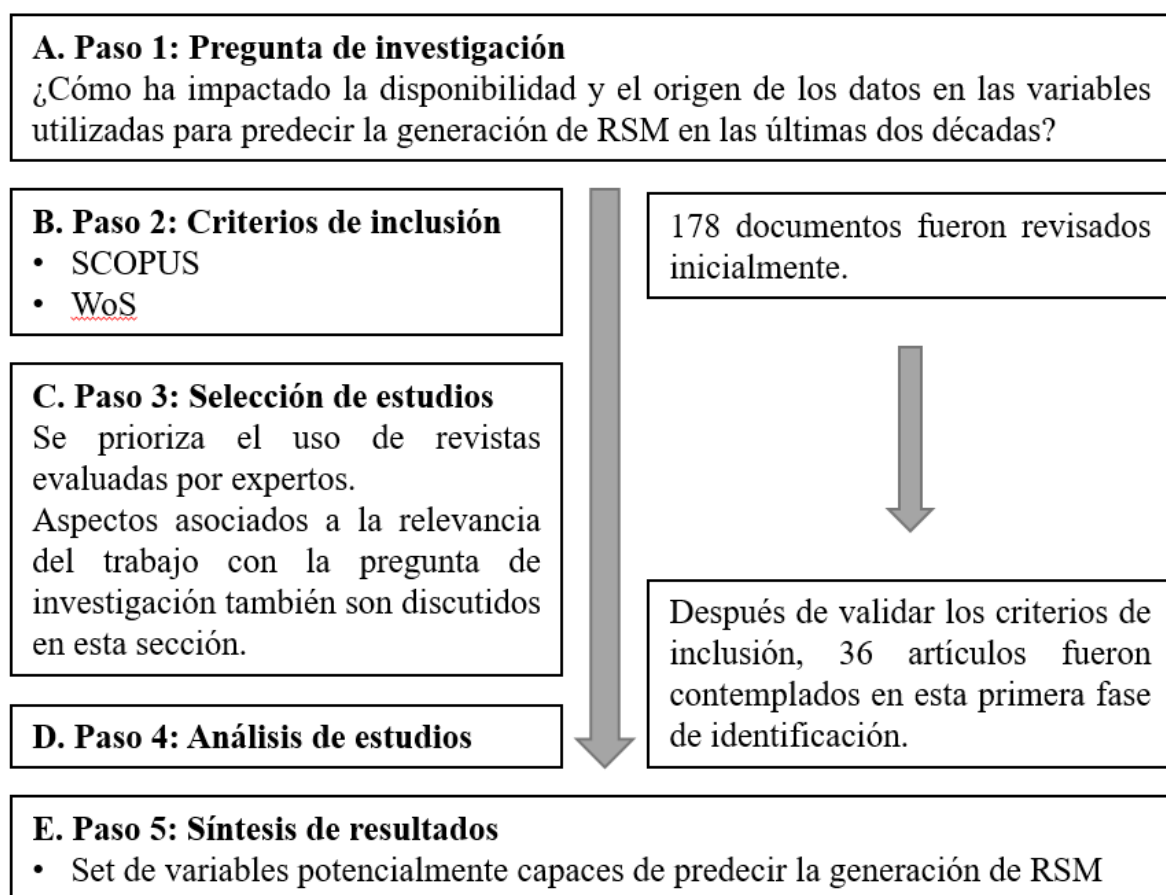


Figura 5 - Flujo metodológico de la revisión literaria realizada sobre la influencia de variables predictoras en la generación de RSM.

4.1.2. Revisión semisistemática: Criterios de inclusión

El proceso de revisión de la literatura se basó en la búsqueda de artículos científicos indexados escritos en inglés que hayan pasado por un proceso de revisión por pares. Se utilizaron los motores de búsqueda Web of Science (WoS) (Thomson Reuters, 2024) y Scopus (Elsevier,

2024). La revisión de la literatura se llevó a cabo mediante la utilización de las combinaciones de palabras clave detalladas en la Tabla 1.

Tabla 1 - Palabras clave utilizadas durante la revisión bibliográfica sobre la influencia de las variables en el pronóstico de la generación de RSM.

Combinaciones de palabras clave
Ciencia de datos + Residuos sólidos municipales + Pronóstico
Aprendizaje automático + Residuos sólidos municipales + Generación
Pronóstico + Residuos sólidos municipales + Acumulación
Variables explicativas + Residuos sólidos municipales + Generación

En los últimos años, un número creciente de investigadores se ha centrado en explorar los beneficios de técnicas predictivas de RSM acompañadas por un desarrollo paralelo en el procesamiento y análisis de datos (A. Xu et al., 2021). Sin embargo, a pesar del avance en los algoritmos de predicción, las variables predictoras parecen haber permanecido constantes y que comparten algunas fuentes comunes de información. En ese sentido, la caracterización de los estudios revisados se llevó a cabo principalmente a través de los siguientes criterios: descripción y agrupación de las variables predictoras involucradas (i.e., socioculturales, ambientales y económicas), objetivos y alcance, fuentes de información y principales hallazgos.

En primer lugar, se identificaron y categorizaron las variables utilizadas como descriptores y se agruparon de manera apropiada. Este paso tiene como objetivo asegurar que todos los artículos estén relacionados con el objetivo de la investigación. Además, este criterio permite sistematizar las variables más comunes para cada dimensión de interés para futuras inferencias en cuanto a influencia en la predicción de RSM refiere, ya que el objetivo de este estudio es proponer mejoras en la identificación de este conjunto de predictores. El segundo criterio se refiere a la relevancia entre el propósito del estudio analizado y el objetivo de la investigación; es decir, validar si este estudio contribuye de manera significativa al análisis de la influencia

del conjunto de predictores en el pronóstico de la generación de RSM para una escala de tiempo dada. El tercer criterio de evaluación fue el tratamiento de datos; es decir, reconocer la relevancia y validez de la información. Este criterio permite comprender y reconocer el vínculo causal entre las fuentes de información y el conjunto de predictores. También muestra las oportunidades de mejora en ausencia o falta de fiabilidad de la información disponible. El cuarto criterio se refiere a la gestión de datos, es decir, la técnica utilizada para procesar la información. Esta sección también permite identificar qué métodos predictivos (p. ej., regresión múltiple, árboles de decisión, series de tiempo, redes neuronales; entre otros) para la generación de RSM se han utilizado más comúnmente con este fin durante los últimos 22 años.

4.1.3. Revisión semisistemática: Estrategia de búsqueda y esquema de revisión

La búsqueda inicial arrojó más de 150 coincidencias y los resultados obtenidos se sometieron a una serie de validaciones a través de criterios definidos en la investigación. En primer lugar, se examinaron los títulos para determinar la relevancia del tema en relación con el alcance del análisis actual. Después, se revisaron los resúmenes y solo se consideraron aquellos que emplearon al menos una de las tres dimensiones mencionadas anteriormente para la siguiente ronda. En un paso siguiente, se seleccionaron estudios de investigación que proporcionaron una descripción de las variables aplicadas, así como detalles de sus fuentes de datos y enfoque metodológico. Una vez finalizados todos estos pasos, se evaluaron en profundidad un total de 36 artículos con fines de este estudio. La frecuencia de las palabras clave de la literatura revisada se presenta en la Figura 6.

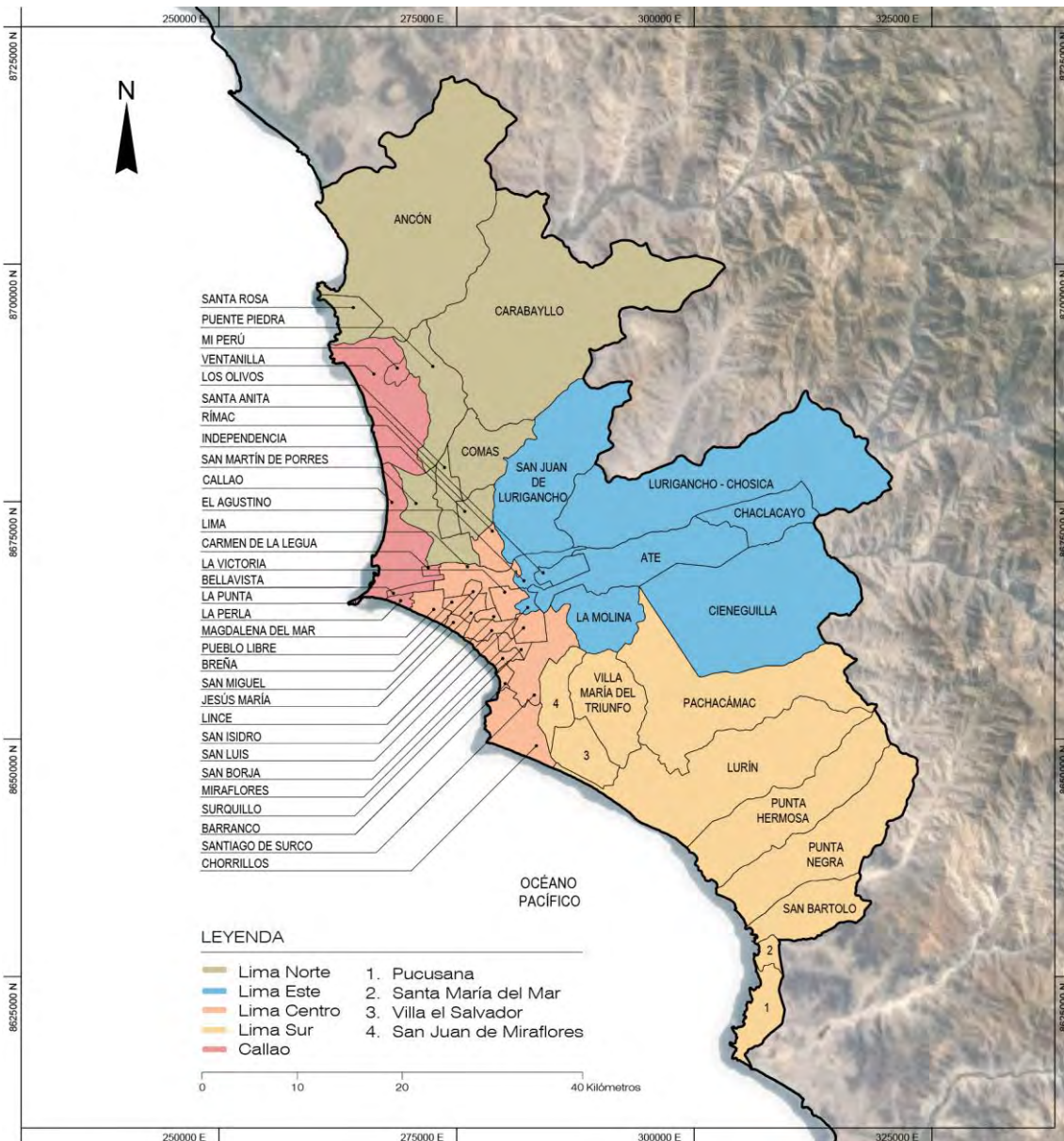


Figura 7 - Mapa de ubicación y conformación distrital del área de estudio: Lima, Perú. La coloración de los distritos corresponde a su ubicación relativa dentro de la Provincia de Lima & Provincia Constitucional del Callao (i.e., Lima Norte, Lima Sur, Lima Centro, Lima Este y Callao).

4.3 Vínculo entre la generación de RSM y el consumo de combustible

4.3.1. Recopilación de información

Con el fin de comprender las variables determinantes en la generación de RSM, se procedió a examinar diversas bases de datos, tanto gubernamentales como no gubernamentales. Estos datos, a pesar de la incertidumbre inherente a sus procesos de construcción y obtención, se identificaron como los más idóneos en términos de disponibilidad y calidad. Para cuantificar

la generación de RSM, se recurrió al sitio web del Sistema de Información de Gestión de Residuos Sólidos (SIGERSOL) como fuente primaria (MINAM, 2019b). Por otro lado, la información obtenida de OSINERMINING permitió recabar datos sobre el uso de combustibles para cocinar y los gastos relacionados con el ámbito doméstico (OSINERGMIN, 2020b). De la misma manera, las características socioeconómicas de la población fueron adquiridas a través del último censo nacional (INEI, 2017). Es relevante destacar que se eligió el año 2019 como el punto de referencia para el análisis. A pesar de contar con datos correspondientes a los años 2020, 2021 y 2022, se optó por no considerarlos debido a las significativas interrupciones en el sistema de gestión de residuos causadas por la pandemia de la COVID-19.

Considerando que la información asociada a la generación de RSM está disponible tanto en SIGERSOL como en los reportes anualizados del MINAM, esta investigación utilizó los datos provistos en la plataforma virtual SIGERSOL (MINAM, 2019b), la cual consolida informes presentados por municipios. Esta elección se fundamenta en la disponibilidad de datos más detallados y desglosados en relación a la generación de RSM, así como en la observación de una menor cantidad de omisiones en comparación con otras fuentes. De hecho, esta percepción fue corroborada durante consultas con el personal del Ministerio del Ambiente, quienes también indicaron que SIGERSOL era la base de datos más adecuada en Perú para extraer datos relacionados con residuos (Verónica Mendoza-Díaz, División de Gestión de Residuos Sólidos, Ministerio del Ambiente, Perú, comunicación personal, abril de 2023).

4.3.2. Análisis exploratorio de las variables seleccionadas

Tras procesar la información recopilada, la selección de variables para el análisis de la generación de RSM debe basarse en la robustez y validez de cada indicador. Esta selección garantiza una base sólida para la toma de decisiones y la formulación de políticas eficaces. Por ello, se deben elegir variables que abarquen diversas dimensiones influyentes en la generación de RSM, incluyendo aspectos sociales, económicos, ambientales y específicamente

relacionados con los residuos. Por otro lado, las variables deben ser capaces de proporcionar información útil y relevante para comprender y abordar eficazmente el problema de la generación de residuos sólidos alimenticios. Dicho de otro modo, las variables seleccionadas deben ser capaces de proporcionar una comprensión más profunda de los patrones y las tendencias en la generación de residuos alimenticios en hogares residenciales urbanos. Para una mejor exploración e interpretación de las consideraciones realizadas respecto a los datos recopilados, la Tabla 2 muestra las relaciones identificadas y las variables seleccionadas.

Tabla 2 - Análisis exploratorio de las variables identificadas como impulsoras de la generación de residuos sólidos alimenticios.

Dimensión	Variable	Descriptor	Fuente de información	Consideraciones
Social / Demográfico	Uso del suelo	Residencial Comercial Industrial	Población (2019) Uso del suelo (2017) Preferencia de cocina (2017) Fuente: (INEI, 2017)	El uso del suelo y las preferencias de cocina de cada vivienda permanecen constantes durante el periodo de estudio.
Económico	Aspectos socioeconómicos	Estatus socioeconómico Gasto en alimentación	Gasto en alimentos para el consumo (2017) Nivel socioeconómico (2019) Fuente: (INEI, 2017)	El gasto medio mensual en alimentación se mantiene constante.
Ambiente	Consumo de combustible exclusivamente para cocinar	Electricidad GLP Gas natural	Aspectos asociados al uso y consumo de combustible para cocinar (2019) Fuente: (OSINERGMIN, 2019)	-
Residuos	Generación de residuos alimenticios per cápita	Generación de residuos alimenticios domiciliarios	Generación de residuos alimenticios domiciliarios por distrito (2019) Fuente: (MINAM, 2019b)	Basado en informes municipales registrados en SIGERSOL

Es importante destacar que diferentes estudios han informado resultados diversos en cuanto a la generación de residuos alimentarios domésticos; sin embargo, es fundamental tener en cuenta que esta variabilidad se deriva de la naturaleza epistémica de los datos y las

circunstancias específicas de cada estudio (Mu'azu et al., 2019). Por ejemplo, en una economía de nivel medio-alto, la proporción de residuos sólidos alimentarios domésticos en relación con el total de RSM suele situarse alrededor del 54 %, coincidiendo con el promedio latinoamericano de aproximadamente el 52 % (Kaza et al., 2018). Esto se asemeja estrechamente al rango identificado en el presente estudio, que se sitúa en el 50 %. Además, algunas estimaciones indican que los consumidores en Perú generan un promedio de 67 kg de residuos alimentarios por persona solamente en la etapa de consumo (Bedoya-Perales & Magro, 2021). En contraste, Vázquez-Rowe et al. (2021) reportan una cantidad diferente, señalando 29 kg por persona, mientras que según el Programa de las Naciones Unidas para el Medio Ambiente (PNUMA), cada hogar produce aproximadamente 72 kg de residuos sólidos alimentarios (UNEP, 2021).

Por otro lado, para comprender las principales relaciones entre las variables seleccionadas como posibles variables predictoras de la generación de residuos alimentarios domésticos, es necesario explorar qué asociaciones se pueden identificar. Dado que el objetivo principal en esta etapa es explorar tanto los parámetros como el resultado final, se realizó un análisis de inferencia causal teniendo en cuenta que, si no se encuentra asociación, las variables no son atribuibles al resultado (Frank et al., 2022). Esto implica que la generación de residuos alimentarios domésticos puede expresarse de la siguiente manera:

$$Y \rightarrow F(\text{consumo de combustible}) + F(\text{uso de suelo}) + F(\text{aspectos socioeconómicos}) + \varepsilon$$

donde el término $F(x)$ representa la variable de entrada que determina el comportamiento y los resultados de la función, mientras que el término Y (i.e., generación de residuos alimentarios domésticos per cápita) denota la variable de salida que describe los valores correspondientes generados por la función en relación con la entrada $F(x)$. Además, ε representa el ruido asociado con la complejidad e incertidumbre del análisis.

4.3.3. Nexa entre la generación de residuos sólidos domiciliarios y el consumo de combustible

Con el fin de entender las variables influyentes en la generación de residuos de alimentos a nivel doméstico, se emplearon datos de OSINERMING para recopilar información sobre el uso de combustibles para cocinar y el gasto en el hogar (OSINERGMIN, 2020a). Es importante destacar que nuestra atención se centró específicamente en el consumo global de combustibles para cocinar (i.e., GN, GLP y electricidad), evitando evaluar la eficiencia de cada uno de estos en el proceso de cocina. Esta decisión se basa en dos consideraciones. Por un lado, nuestro interés principal radica en comprender el consumo total de combustibles utilizados para cocinar en lugar de la eficiencia de fuentes individuales. Por otro lado, la preferencia de combustible para cocinar dependerá no sólo del poder adquisitivo, sino también de la disponibilidad de combustible. Finalmente, debido a la necesidad de gestionar datos a un nivel de observación más preciso (i.e., por distrito o bloque), los datos mencionados anteriormente se complementaron con información del área de interés a través de informes locales y municipales. Todos los datos fueron procesados posteriormente en MS Excel y Stata.

4.3.4. Análisis de regresión lineal múltiple

Mediante un análisis de regresión, es posible observar la relación entre dos o más variables a través de ecuaciones (Abuín, 2008). Se utilizó una regresión lineal para cuantificar la relación entre una variable predictora y una variable de respuesta. Esta técnica estadística puede representarse como en la Figura 8. Este modelo tiene la forma:

$$Y = B_0 + B_1X_1 + \dots B_mX_m + \varepsilon$$

donde Y representa la variable dependiente o variable respuesta; X_i constituyen las variables explicativas o regresoras; B_j son los parámetros del modelo y los que miden la influencia que las variables explicativas tienen sobre las regresoras, y ε es la parte aleatoria (Véliz, 2023).

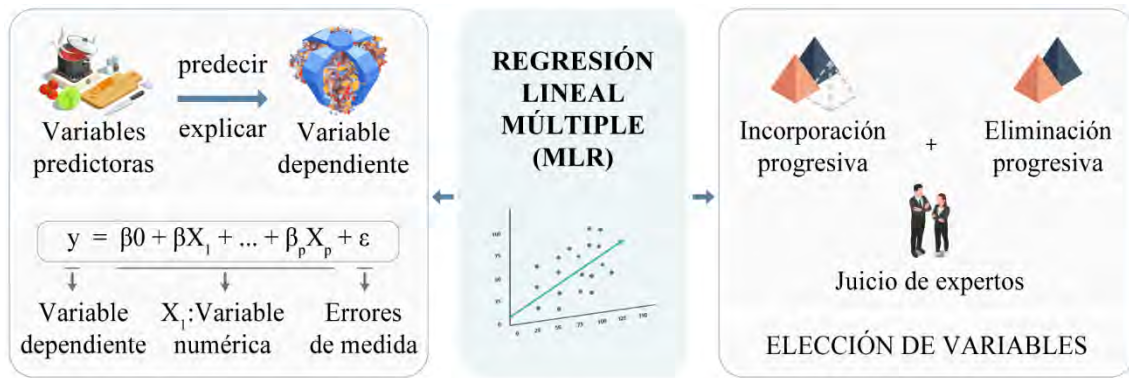


Figura 8 - Representación esquemática y conceptual de una regresión lineal múltiple.

4.3.5. Análisis de regresión lineal múltiple: Significancia global e individual de los coeficientes del modelo

Debido a la naturaleza del modelo, es posible establecer la significancia global de los coeficientes del modelo. Esta significancia puede analizarse evaluando la hipótesis nula: $H_0: B_0 = B_1 = B_2 \dots = B_m = 0$ versus la hipótesis alternativa: H_A : Alguno de los B_i es igual a 0. Esto quiere decir que, si la hipótesis nula se rechaza, se concluirá que el modelo puede usarse para la predicción (Véliz, 2023).

En esta investigación, la generación domiciliar de residuos de alimentos per cápita (GPC_HFW) fue definida como la variable dependiente. Además, se propusieron tres enfoques analíticos para analizar las particularidades de la región de interés. El primer enfoque analiza todas las observaciones posibles considerando el criterio de combustible predominante (i.e., GLP) y electricidad. El segundo enfoque tiene como objetivo analizar solo las observaciones que utilizan los tres tipos de combustibles analizados. Finalmente, el último enfoque relaciona las observaciones con una variable construida como una función del uso de combustible (i.e., $GLP + GN$), considerando sus distribuciones y usos en todo el territorio bajo estudio.

4.4 Implementación del modelo predictivo: *random forest*

La selección del modelo predictivo adecuado dependerá de varios factores, como la naturaleza de los datos disponibles, el tamaño del conjunto de datos y el objetivo de la predicción, entre

otros aspectos específicos del problema (Sunayana et al., 2021). Por lo tanto, es fundamental comprender las ventajas y desventajas de cada tipo de modelo. Así pues, es esencial distinguir entre dos enfoques generales en el aprendizaje automático: el aprendizaje supervisado y el no supervisado (Munir et al., 2023). La diferencia clave entre ambos radica en que el aprendizaje supervisado se enfoca en la predicción utilizando datos que están etiquetados previamente, mientras que el aprendizaje no supervisado se dedica a descubrir patrones y estructuras en datos que no cuentan con etiquetas (Guo et al., 2021).

De esta manera, entre los modelos de aprendizaje supervisado más comúnmente empleados para predecir la generación de RSM, se destaca el uso del *random forest* debido a su versatilidad para manejar tanto datos categóricos como numéricos. Este modelo demuestra eficacia al predecir relaciones no lineales y presenta una robustez frente a valores atípicos; no obstante, es importante considerar que puede demandar recursos computacionales significativos (Kannangara et al., 2018). Por otro lado, es importante mencionar que las redes neuronales artificiales (conocidas como ANN, por sus siglas en inglés) representan otro enfoque ampliamente utilizado, ya que son capaces de modelar relaciones complejas. Sin embargo, es relevante tener en cuenta que las ANN suelen requerir conjuntos de datos extensos para su entrenamiento y pueden estar sujetas al riesgo de sobreajuste (M A Abdoli et al., 2011).

En ese sentido, el *random forest* surge de la combinación de múltiples árboles de decisión. La función primordial de este tipo de árbol es facilitar la predicción de un resultado mediante el recorrido de los diferentes niveles jerárquicos del árbol. La Figura 9 presenta un esquema de árbol de decisión, en el cual cada nodo experimenta una bifurcación basada en un criterio de división específico. Adicionalmente, estos árboles son conocidos por su resistencia frente a valores atípicos, atribuida a la estructura jerárquica de las divisiones, y su invariabilidad ante transformaciones monótonas de las variables (Cutler et al., 2012). Finalmente, el clasificador

final que se obtiene es el que asigna la categoría con mayor votación en el caso de la clasificación y, el promedio de predicciones, en el caso de la regresión (Véliz, 2023).

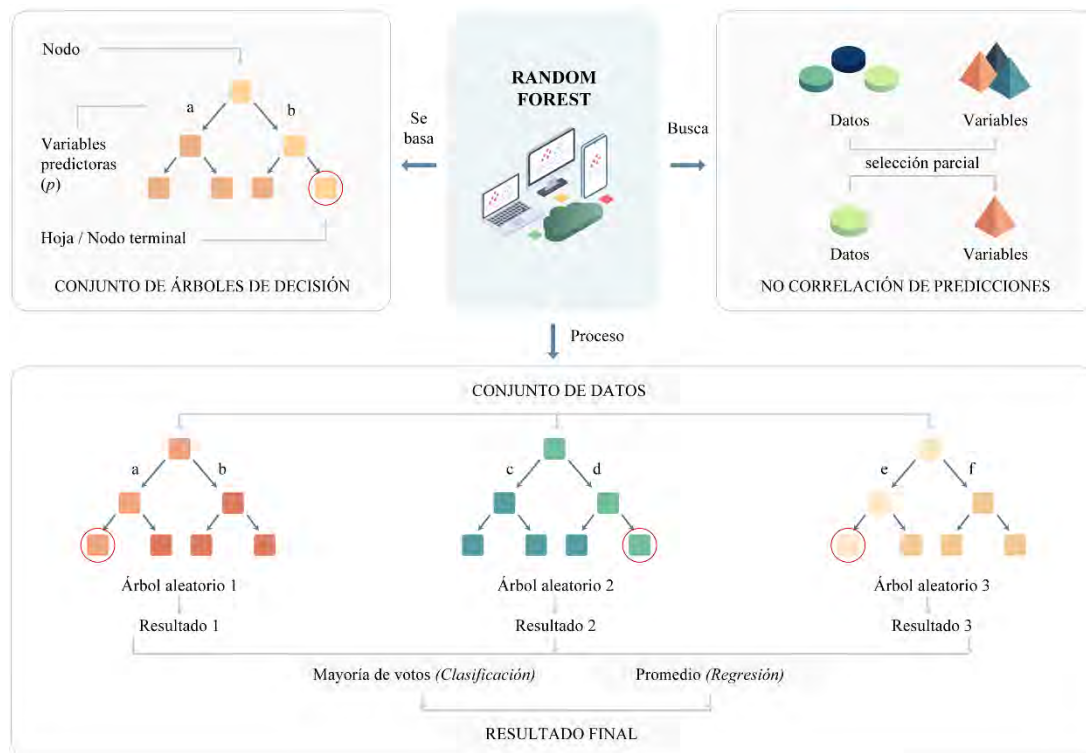


Figura 9 - Representación conceptual del método random forest.

4.4.1. Recolección de datos

En el proceso de recolección de datos para la presente tesis, se emplearon dos fuentes fundamentales: la plataforma web del Sistema de Información de Gestión de Residuos Sólidos (SIGERSOL) (MINAM, 2019b) y la Encuesta Residencial de Consumo y Uso de Energía (ERCUE) (OSINERGMIN, 2021a) correspondiente al año 2019. La combinación de estas herramientas permitió obtener una amplia gama de información, destacando la diversidad de datos nominales y categóricos recopilados. La plataforma SIGERSOL provee la información de la variable dependiente; es decir, el valor de la generación per cápita a nivel distrital, por lo que el número de observaciones es de 50 (i.e., 43 distritos correspondientes a la Provincia de Lima y 7 distritos correspondientes a la Provincia Constitucional del Callao). Por otro lado, la encuesta ERCUE, aplicada a una muestra de 1797 participantes, proporcionó una visión

detallada de las percepciones y comportamientos de los encuestados en relación con las características del hogar (p. ej., edad, nivel de educación), consumo de combustible (p. ej., preferencias en el tipo de combustible para cocinar y gasto mensual asociado en soles) y características económicas (p. ej., gasto de la familia anualizado). Esto permitirá una interpretación más completa y precisa de los resultados, contribuyendo así al logro de los objetivos planteados en la investigación. La Tabla 3 muestra los datos provistos por ERCUE que se utilizarán como variables independientes. En cuanto a la variable dependiente, esta se construirá en función de la información asociada a la generación per cápita de residuo sólido municipal a nivel distrital (*GPC_MSW_MINAM*) facilitada por el MINAM.

Adicionalmente, debido a la complejidad de ciertas variables que incorporaban más de cuatro descriptores, se generó variables adicionales con el fin de condensar y agrupar las características inherentes a la información. Una vez concluido el análisis exploratorio y realizado el adecuado preprocesamiento, se procedió a la implementación del modelo. Posteriormente, se llevó a cabo una fase de optimización que abarcó la selección de los hiperparámetros más idóneos, así como la elección definitiva del conjunto final de variables predictoras. Este análisis garantiza la robustez y eficiencia del modelo, contribuyendo a la calidad de los resultados obtenidos en el estudio.

Tabla 3 - Variables independientes utilizadas en la predicción de residuos sólidos municipales. Fuente: Encuesta Residencial de Consumo y Usos de Energía (ERCUE) 2019-2020 (OSINERGMIN, 2021a).

Variables independientes	Descripción de la variable
Electricity_cook	Indica si en la vivienda se usa la electricidad para la cocción de alimentos
LPG_cook	Indica si en la vivienda se usa el GLP para la cocción de alimentos
NG_cook	Indica si en la vivienda se usa el GN para la cocción de alimentos
Electricity_access	Indica si dentro del hogar se cuenta con acceso a servicio de Electricidad
NG_access	Indica si dentro del hogar se cuenta con acceso a servicio de Gas Natural

LPG_access	Indica si dentro del hogar se cuenta con el servicio de GLP
Monthly_electricity_payment	Pago mensual en soles S/ (último mes) por consumo de Electricidad
Monthly_NG_payment	Pago mensual en soles S/ (último mes) por consumo de Gas Natural
Monthly_LPG_payment	Pago mensual en soles S/ (último mes) por consumo de GLP
Price_LPG_10kg	¿Cuál fue el precio de la última adquisición? – Balón de GLP 10kg
Total_annual_expense	Gasto anual total en soles (S/)
Number_of_HM	Número de miembros del hogar
Average_kWh_month	Consumo promedio mensual por hogar (kWh-mes)
Edad	Edad
Nivel de educación	Grado de nivel de educación alcanzado

4.4.2. Preprocesamiento de la información

La etapa de preprocesamiento de la información se enfocó principalmente en la homogenización de la base de datos obtenida de fuentes gubernamentales a nivel de vivienda, específicamente para el caso de estudio. Para lograr este objetivo, se llevó a cabo un análisis exploratorio de las variables, dirigido a identificar posibles valores atípicos y datos faltantes. La detección de valores atípicos desempeña un papel esencial en la validación y limpieza de datos para garantizar la calidad y fiabilidad de los resultados obtenidos. En este estudio, se implementó el método de *box plot*, una herramienta gráfica basada en percentiles, para identificar posibles valores atípicos en las variables analizadas.

El *box plot* proporciona una representación visual clara de la distribución de los datos, destacando áreas específicas donde podrían existir observaciones inusuales. Mediante la delimitación de los cuartiles y la identificación de los llamados "bigotes", este enfoque permite identificar valores que se encuentran significativamente fuera de la tendencia general de la muestra. La utilización de percentiles facilita una evaluación más precisa de la dispersión de

los datos, ayudando a distinguir entre variabilidad normal y observaciones potencialmente anómalas. Este proceso de detección de valores atípicos no solo contribuye a la integridad de los resultados, sino que también permite una interpretación más robusta de los patrones y tendencias presentes en los conjuntos de datos analizados (Véliz, 2023).

4.4.3. Ingeniería de variables

En la fase de ingeniería de variables, se llevó a cabo la construcción de variables adicionales en función de las originales con el objetivo de mejorar la representación de los datos y potenciar la capacidad predictiva del modelo. Este proceso se basó en una evaluación de la interacción entre las distintas variables, considerando la posibilidad de descubrir patrones emergentes y relaciones no evidentes en los datos originales. En ese sentido, considerando que los datos originales proveían del gasto del último mes, era factible cuantificar el consumo en metros cúbicos equivalente de combustible en función del precio del combustible del año de referencia. Adicionalmente, debido a las políticas de distribución del GN, las variables asociadas a esta y GLP fueron consideradas como uno. En suma, se hicieron las siguientes relaciones, considerando las condiciones del mercado (i.e., precio de m³ de GN) y equivalencias volumétricas de GN y GLP (i.e., cantidad de m³ equivalente de GLP). Donde, las variables *'Total_gas'* y *'Total_Pay'* hacen referencia a la unificación del consumo de combustible de GN + GLP y al costo mensual en soles generado por dicho consumo, respectivamente.

$$'GN'_{m3} = 'Monthly_NG_payment' / precio_por_{m3_NG}$$

$$'GLP'_{m3} = 'Monthly_LPG_payment' * cantidad_{m3_{10kg}} / 'Price_LPG'_{10kg}$$

$$'Total_gas' = 'GLP'_{m3} + 'GN'_{m3}$$

$$'Total_Pay' = 'Monthly_NG_payment' + 'Monthly_LPG_payment'$$

Asimismo, se construyó la variable objetivo *'RSM_VIVIENDA'* en función de la generación per cápita de residuo sólido municipal a nivel distrital (*GPC_MSW_MINAM*) y el número de

habitantes del hogar (*Number_of_HM*). Esto con la finalidad de obtener mediciones en una misma escala de observación (i.e., nivel vivienda). Para ello, se utilizó la siguiente expresión:

$$'RSM_VIVIENDA' = 'GPC_MSW_MINAM' * 'Number_of_HM'$$

4.4.4. Calibración de parámetros del modelo

Para esta etapa, se implementó una búsqueda de parámetros utilizando validación cruzada por *fold*. La división de datos se llevó asignando el 70 % de los datos para el conjunto de entrenamiento y reservando el 30 % restante para el conjunto de testeo o validación. Esta división se ejecutó con el propósito de evitar problemas de sobreajuste o subajuste, garantizando que el modelo se entrenara y evaluara en conjuntos de datos independientes. Posteriormente, sobre los datos de entrenamiento, se realizó una búsqueda grillada de parámetros mediante técnicas como *GridSearchCV*. Esta metodología permitió explorar sistemáticamente un conjunto predefinido de combinaciones de parámetros, evaluando el rendimiento del modelo con validación cruzada en cada iteración; esto con la finalidad de encontrar el equilibrio óptimo que maximizara la generalización del modelo a datos no vistos.

4.4.5. Indicadores de performance del modelo

Para evaluar el modelo de predicción de *random forest*, se seleccionaron dos indicadores de evaluación: el error cuadrático medio (MSE) y el coeficiente de determinación (R^2). El MSE representa la discrepancia entre los valores observados y reales. Cuanto más cercano sea el valor de MSE a 0, mayor será la precisión del modelo. El R^2 refleja el grado de ajuste entre las variables y, cuanto más cercano sea a 1, mejor será el ajuste. Estas métricas vienen dadas por las siguientes expresiones:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \qquad MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

Donde, n es el número de observaciones, \hat{Y}_i es el valor predicho por el modelo, Y_i es el valor observado, \bar{Y} es el valor medio de la cantidad de residuos sólidos domiciliarios generados por vivienda.

CAPÍTULO 5: ANÁLISIS Y DISCUSIÓN DE RESULTADOS

En esta sección, se detallan los hallazgos clave de la investigación y se demuestra la aplicación del marco metodológico presentado en el capítulo anterior en el contexto del caso de estudio. En primer lugar, se presentan los resultados de la revisión literaria y del análisis exploratorio del caso de estudio realizado. A continuación, se analizan los hallazgos relativos a la identificación de variables predictoras y su relación con la generación de RSM, con énfasis en el residuo sólido alimenticio en hogares y el consumo de combustible para cocinar. Finalmente, se describen los resultados de la implementación del modelo *random forest* para la predicción de la generación de RSM a nivel vivienda.

5.1 Resultados derivados de la revisión semisistemática

La revisión bibliográfica ha permitido explorar una amplia variedad de fuentes académicas y estudios previos relacionados con la gestión de residuos sólidos a nivel municipal. El objetivo principal de esta sección es resumir y analizar las variables más destacadas identificadas en la literatura, que desempeñan un papel significativo en la generación de residuos sólidos en entornos urbanos. Estos hallazgos sientan las bases para un análisis más profundo y la formulación de modelos predictivos efectivos en el contexto de esta investigación. Por consiguiente, del análisis realizado se desprende que la generación de RSM depende —en gran medida— de las actividades humanas en la tecnosfera y de cómo se disponen estos residuos, afectando —de algún modo— al ambiente. Por lo tanto, se presenta una dimensión sociocultural, económica y ambiental como una alternativa potencialmente capaz de prever la generación de RSM en una escala de tiempo determinada para una comunidad con características previamente definidas.

5.1.1. Consideraciones asociadas a variables socioculturales

Las consideraciones socioculturales en la generación de RSM llevan consigo múltiples suposiciones y generalizaciones basadas en constructos socialmente aceptados que deben ser

tomadas en consideración (S. Khan et al., 2022). Por ejemplo, se deben considerar aspectos relacionados con la actitud de la población local hacia los residuos, la relación entre la gobernanza y la estabilidad política (Kundariya et al., 2021). Por lo tanto, esta sección explora los principales descriptores socioculturales (i.e., edad, condiciones demográficas y educación) que permiten una interpretación y caracterización sencilla de la comunidad estudiada.

Consideramos que esta dimensión es la que caracterizará a la población en análisis en cuanto a su comportamiento frente a la generación de RSM. En esa línea, Sinha y Prabhudev (2016) señalan que los grupos de población más jóvenes (menores de 25 años) contribuyen de manera diferente a la generación de RSM en comparación con las personas mayores. Sin embargo, Buenrostro et al. (2001) sostienen que esta variable por sí sola no muestra un potencial predictivo como tal, sino que se utiliza para agrupar a la población en estudio. También mencionan que, según la técnica de manejo de datos aplicada, la edad presenta correlaciones significativas con respecto a la generación de RSM. Por ejemplo, la investigación realizada por Lebersorger & Beigl (2011) encontró correlaciones positivas entre la edad y RSM: 0.70 con una significancia de 0.001.

Variables como el crecimiento poblacional, la densidad de población por metro cuadrado y el número de hogares por vivienda muestran una correlación positiva con la generación de RSM (Lebersorger & Beigl, 2011). Las dos primeras variables son relevantes debido a su fluctuación a lo largo del tiempo, lo que dificulta las proyecciones (Sunayana et al., 2021); mientras que la última (i.e., número de hogares por vivienda) está estrechamente relacionada con las primeras. A mayor cantidad de hogares, se espera que sea mayor la generación de RSM (Dyson & Chang, 2005). Otro aspecto relevante a considerar para el uso de estas variables es la disponibilidad de información, ya que la fuente de información variará según el objeto de estudio. Por ejemplo, si se mide el crecimiento general de la población, los censos nacionales son bastante útiles; sin embargo, si se requiere el número de habitantes por hogar, es más recomendable utilizar

cuestionarios debido a su capacidad para recopilar datos detallados y específicos mediante preguntas dirigidas.

En términos de educación, a partir de la revisión de Kolekar et al. (2016) y los estudios de Keser et al. (2012) y Grazhdani (2016), se determina que la educación influye significativamente en la generación de RSM, especialmente en países en desarrollo. Por otro lado, también se han tenido en cuenta variables asociadas con la raza, el género o la migración, debido a la importancia de reconocer y aceptar las diversas representaciones culturales con respecto a la generación de RSM (Mohd Yusof et al., 2012). La información proveniente de este tipo de variables permite no solo obtener modelos más eficientes al tomar decisiones relacionadas con la generación de RSM (Z. S. Li et al., 2011); sino también permite visualizar la heterogeneidad de orígenes en el mismo sector, demostrando la necesidad de un análisis transversal. Esto debido a que algunas variables culturales y prácticas sociales influyen en la generación y gestión de RSM (Albert & Olutayo, 2021). Por ejemplo, según Sinha y Prabhudev (2016), en la India, las mujeres poseen un entendimiento más profundo del conocimiento de la gestión de residuos domésticos que los hombres debido a su participación en las actividades domésticas.

Teniendo en cuenta todo este contexto, las variables más comúnmente utilizadas para caracterizar a una población en términos socioculturales, según la revisión de la literatura, serían la edad, el nivel educativo y el crecimiento poblacional. La principal razón de esto podría ser el hecho de que la información relacionada con estas variables suele estar ampliamente disponible y es mucho más accesible de encontrar que la relacionada con otros tipos de variables mencionadas en esta sección. Además, varios estudios revisados consideran que esta trilogía de variables es suficiente para una primera aproximación; sin embargo, Ali Abdoli et al. (2011) sugieren que la dimensión sociocultural por sí sola no proporciona buenos resultados para predecir la generación de RSM, porque no tiene en cuenta otras categorías de análisis.

Esto es especialmente cierto con respecto a las variables económicas, ya que se espera que el nivel económico de un país impacte sustancialmente en la generación de residuos, ya que a mayor crecimiento económico y consumo de una sociedad, mayor será la generación de RSM (Namlis & Komilis, 2019). Este tipo de información ayuda a quienes toman las decisiones de diseñar planes de gestión según sus escalas territoriales, condiciones demográficas (Karadimas & Orsoni, 2006) y aspectos económicos (Younes et al., 2015).

5.1.2. Consideraciones asociadas a variables ambientales

La generación y acumulación de RSM afecta no solo a la población expuesta (Cervantes et al., 2018), sino también al entorno construido (Zaman, 2014), lo que lo hace relevante en su pronóstico. Por lo tanto, una caracterización de aspectos ambientales, como el consumo de combustible/energía (Chhay et al., 2018) o factores climáticos como la temperatura (Abbasi et al., 2018) y la precipitación (Johnson et al., 2017), es esencial. Estas variables cumplen con la función de representar la relación entre la unidad de análisis con su entorno construido y natural en tareas relacionadas con la generación de RSM (Vu et al., 2019). Por lo tanto, las características ambientales y climáticas son importantes para pronosticar la generación de RSM, ya que algunas suposiciones pueden variar según la temporada o patrones de oscilación térmica (Abbasi et al., 2018).

En cuanto a las variables climáticas, la temperatura se considera la variable más relevante por delante de la precipitación y la humedad (Vu et al., 2019). Según Han et al. (2017), la temperatura tiene una fuerte correlación con la generación de RSM, ya que afecta los patrones de consumo (Keser et al., 2012), especialmente en el uso de carbón, madera, etc. Por esta razón, Abbasi et al. (2018) argumenta que una caracterización en términos estacionales es importante siempre que las características específicas del sitio lo permitan.

En cuanto a las variables ambientales, el consumo de energía está directamente relacionado con el uso de combustibles fósiles (Shah et al., 2021). Así, Cárdenas-Mamani et al. (2022) llevaron a cabo un estudio en Lima (Perú) en el que relacionaron el consumo de energía (i.e., electricidad, GN y GLP) con las actividades diarias del hogar (i.e., cocina, calefacción y sistemas de aire acondicionado, entre otros), evidenciando la conexión entre el estatus económico y el nivel de consumo de energía (i.e., los distritos de ingresos altos tienen un mayor consumo promedio de energía).

5.1.3. Consideraciones asociadas a variables económicas

Las variables económicas representan un aspecto altamente relevante en el diseño, planificación y ejecución de actividades relacionadas con la gestión de RSM. Según Kaza et al. (2018), los países de ingresos altos muestran una gestión más eficiente de RSM que los países de ingresos intermedios o bajos. Esta afirmación también implica que en los países de bajos ingresos existe una notable disparidad entre los sectores urbanos y rurales, favoreciendo al primero. Por ejemplo, en términos de recolección de residuos sólidos, en los países de bajos ingresos, los servicios de recolección de residuos son un paso crítico para reducir la contaminación y mejorar la salud humana.

Asimismo, resulta relevante analizar la influencia del sector económico, principalmente en los patrones de consumo (Vitorino de Souza Melaré et al., 2017). Estos patrones de consumo pueden describirse a través del crecimiento económico, procesos industriales, actividades económicas, entre otros (Namlis & Komilis, 2019). Sin embargo, para Gu et al. (2016), este desarrollo económico no ha sido acompañado por una acción estratégica robusta de gestión de RSM aumentando la complejidad de análisis (Abbasi & Hanandeh, 2016). Además, esta complejidad también se explica en la sinergia de esta dimensión económica con las otras categorías estudiadas anteriormente.

Bajo esta perspectiva, las variables más relevantes estudiadas en esta dimensión son las siguientes: producto interno bruto per cápita (Beigl et al., 2008), ingreso per cápita (Nguyen et al., 2021) y el producto nacional bruto (M. Khan & Burney, 1989). Cada descriptor en esta dimensión se enfoca en medir algún indicador económico y su uso dependerá de la disponibilidad de datos y la escala del estudio. En consecuencia, debido a la accesibilidad de la información, el descriptor relacionado con el ingreso per cápita a menudo se sustituye por el producto interno bruto per cápita (Popli et al., 2021). Finalmente, es importante resaltar que todas las variables descritas anteriormente variarán no solo según el área geográfica, sino también según las características de gestión de cada región.

La Figura 10 resume las tres dimensiones analizadas (i.e., sociocultural, ambiental y económica), seguidas por sus principales descriptores utilizados en la literatura revisada. Esta representación gráfica exhibe la sinergia entre las variables posibles que pueden emplearse en la predicción de los RSM.

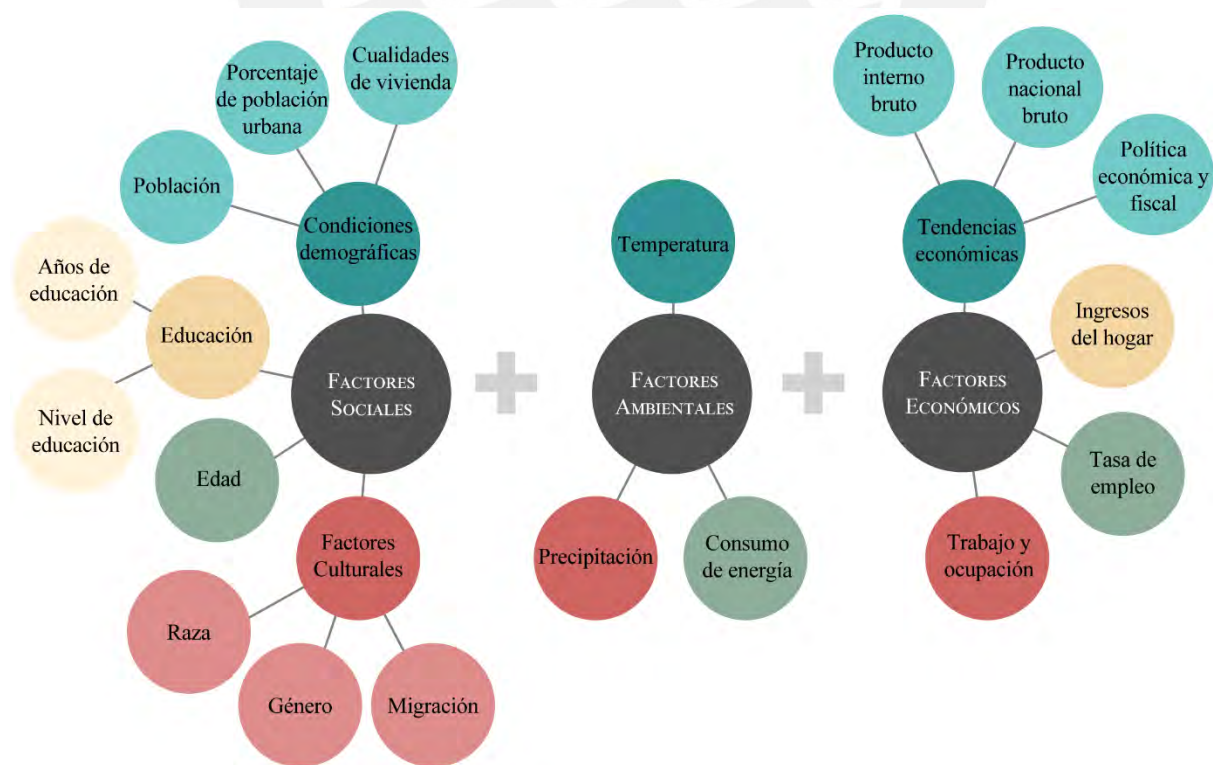


Figura 10 - Principales variables asociadas a la predicción de residuos sólidos.

5.1.4. *Perspectiva holística del pronóstico de RSM*

Como cualquier sistema abierto, las características de la unidad de análisis no están aisladas entre sí, sino que, por el contrario, están interrelacionadas; por lo que es muy importante comprender no solo el funcionamiento individual de cada variable, sino también cómo estas pueden influirse y relacionarse mutuamente (Daly & Farley, 2011). Así, para lograr una representación global y aproximada del fenómeno en estudio, se necesitan generalizaciones sobre las actitudes hacia los RSM (Torrente-Velásquez et al., 2021), generando así una mayor incertidumbre conforme aumenta la complejidad de la relación entre las variables (Munda, 2004). Por ejemplo, según la literatura revisada, la variable de ingresos económicos está estrechamente vinculada al nivel de educación, y este último lleva a entender que las poblaciones con educación secundaria completa o un nivel académico más alto tienden a generar menos residuos (Qu et al., 2009). Sin embargo, es posible que existan ciertas excepciones a esta regla que deben tenerse en cuenta para evitar interpretaciones erróneas o valores atípicos en el procesamiento de la información.

En cuanto a las variables influyentes, sugerimos que las que representan cada dimensión (i.e., sociocultural, ambiental o económico) permitan segmentar la muestra. Para la parte sociocultural, es importante que la población se represente en términos de edad, condiciones demográficas (p. ej., densidad de población), nivel de educación o características culturales. La razón de esto está vinculada al hecho de que cada persona tiene una representación cultural específica. El aspecto ambiental puede representarse mediante variables como humedad, velocidad del viento, precipitación y temperatura, ya que estos afectan los patrones de consumo y estos, a su vez, afectan la generación de RSM. Finalmente, el aspecto económico complementa adecuadamente la representación del estudio de caso, en el cual —dependiendo del objeto de análisis— la variable relacionada con el ingreso económico es la más relevante.

La Tabla 4 muestra las principales variables identificadas en los trabajos revisados. Bajo esta perspectiva, se puede inferir que la generación de RSM es ampliamente influenciada por las actividades humanas y la forma en que se gestionan estos residuos, lo que puede impactar directamente en el ambiente. Asimismo, también se aprecia que los aspectos económicos y socioculturales están estrechamente entrelazados. Por ejemplo, el nivel de ingresos influye en el comportamiento, el estilo de vida y los patrones de consumo (Beigl et al., 2008). En ese sentido, es fundamental seleccionar un conjunto óptimo de variables predictoras que representen con precisión el contexto sociocultural, económico y ambiental del caso de estudio, lo que permitirá una predicción efectiva de la generación de RSM. Asimismo, cada categoría de análisis (i.e., socioculturales, ambientales y económicas) describe una dimensión del estudio de caso y, a su vez, es explicada por distintas variables, cuya aplicación y uso dependerán principalmente de la fuente de información y del alcance del estudio.

Tabla 4 - Variables identificadas como predictoras en el proceso de generación de RSM. Adaptado de Izquierdo-Horna, Kahhat, Vázquez-Rowe, (2022).

Variable predictora	Método de predicción	Autor - año
PBI per cápita Tasa de mortalidad infantil Población urbana envejecida Tamaño de los hogares Esperanza de vida al nacer	Métodos de regresión	(Beigl et al., 2008)
Temperatura Nivel de empleo Renta real de los hogares Población	Métodos de regresión	(M Ali Abdoli et al., 2011)
Salario mensual Personas por vivienda Edad Nivel de estudios Número de empleados Superficie de las instalaciones Número de días laborables Horas de trabajo al día	Métodos de regresión	(Buenrostro et al., 2001)
Género Educación Edad Trabajo y ocupación	Análisis exploratorio	(Sinha & Prabhudev, 2016)

Tamaño de la familia Renta mensual per cápita Nivel de estudios	Métodos de regresión	(Trang et al., 2017)
Género Raza Edad Educación Ingresos Ocupación Características de la vivienda	Análisis exploratorio	(Mohd Yusof et al., 2012)
Cultura urbana Migración Género	Análisis exploratorio	(Albert & Olutayo, 2021)
Población urbana Producto Interior Bruto (PIB) Tasa de alfabetización urbana Incidencia de la pobreza urbana Tamaño de los hogares urbanos Tasa de desempleo urbano	Métodos de regresión	(Popli et al., 2021)
Ingreso medio mensual Consumo medio mensual Población Número de personas Espacio ocupado Superficie total	Bosque aleatorio K-vecinos más cercanos SVM	(Nguyen et al., 2021)
Tamaño total del hogar Nivel de ingresos del hogar Nivel de estudios	Análisis exploratorio	(Kolekar et al., 2016)
Producto interior bruto Consumo interno de materiales Productividad de los recursos	Red neuronal artificial	(Antanasijević et al., 2013)
Edad Ingresos personales Nivel de estudios Tasa de empleo Hogar propio	Árboles de decisión Red neuronal artificial	(Kannangara et al., 2018)
Crecimiento demográfico Renta Nivel de educación Tamaño de los hogares Situación profesional Temperatura Precipitaciones Consumo de energía Niveles de consumo	Red neuronal artificial SVM Análisis de regresión Árboles de decisión	(Abdallah, Talib, et al., 2020)

5.1.5. Revisión semisistemática: Limitaciones e incertidumbres

El nivel de descripción y representación del área de estudio desempeña un papel importante en la elección de variables predictoras, lo que significa que cuanto menor sea el nivel de medición (p. ej., hogar, comunidad, ciudad), mejores resultados se obtendrán en la predicción de RSM con respecto al nivel de observación (p. ej., distrito, provincia, departamento). Por ejemplo, si el análisis de predicción se realiza considerando al país como nivel de observación y a los distritos como nivel de medición, el nivel de incertidumbre tenderá a aumentar debido a la pérdida de control de la información en la construcción de la base de datos. No obstante, con un bajo nivel de medición (i.e., hogar), los resultados de la predicción son potencialmente más precisos debido a una mejor gestión de datos. Comprender el nivel de incertidumbre al estimar la generación de residuos sólidos permitirá conocer el grado de representatividad y variabilidad de la muestra (i.e., incertidumbre epistémica). Además, permitirá estimar el nivel de generalización de los resultados obtenidos. La Figura 11 muestra la variación de la incertidumbre asociada a la estimación de la generación de RSM bajo el escenario de un nivel de observación a nivel de país (eje y) según el nivel de medición (eje x).

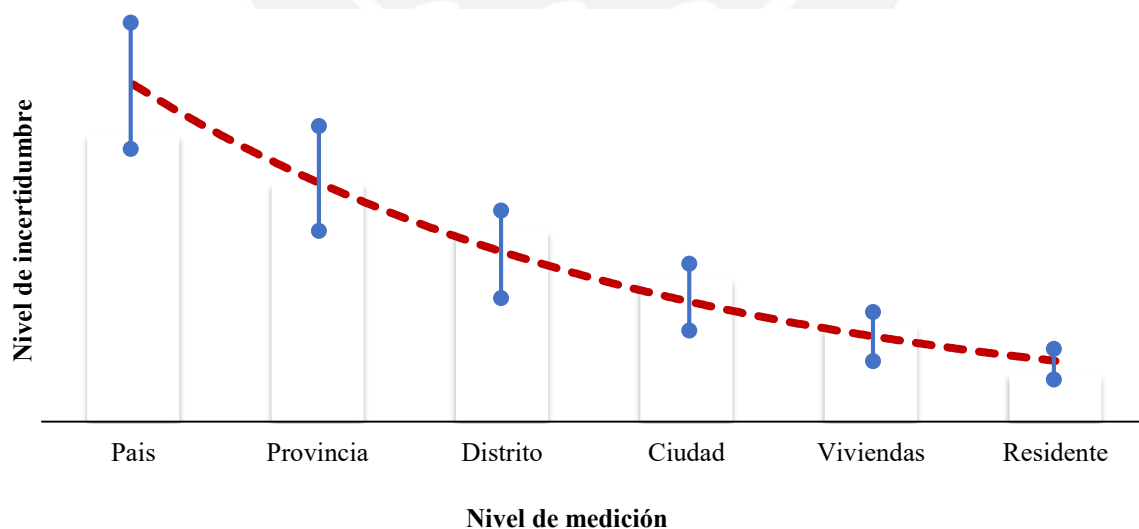


Figura 11 - Variación de la incertidumbre en las estimaciones de RSM según el nivel de medición para un escenario de observación equivalente a país.

5.2 Resultados derivados del análisis contextual del caso de estudio

En términos socioeconómicos, para el año 2019 Lima Metropolitana se dividió en diferentes segmentos, representados por el 5.1 %, 22.8 %, 44.3 %, 22.4 % y 5.5 % para los niveles socioeconómicos (NSE) A, B, C, D y E, respectivamente (APEIM, 2019), lo que refleja una variación significativa en el comportamiento socioeconómico a lo largo de sus 50 distritos. Por ejemplo, en el NSE A, más del 50% de los hogares residen en edificios, y la mayoría utiliza balón de gas para cocinar (más del 85%), mientras que en el NSE E, más del 50% de las viviendas tienen paredes de adobe y alrededor del 41.1% de los hogares emplea leña para cocinar. Además, se observa una clara distinción en los materiales de construcción utilizados en las viviendas entre los diferentes NSE, con el 56.1% de los hogares del NSE C teniendo al concreto como material predominante de sus pisos, en contraste con el NSE E donde el 60.8% tiene como material predominante de sus pisos la tierra. La Figura 12 proporciona una representación visual de la composición socioeconómica de estos distritos.

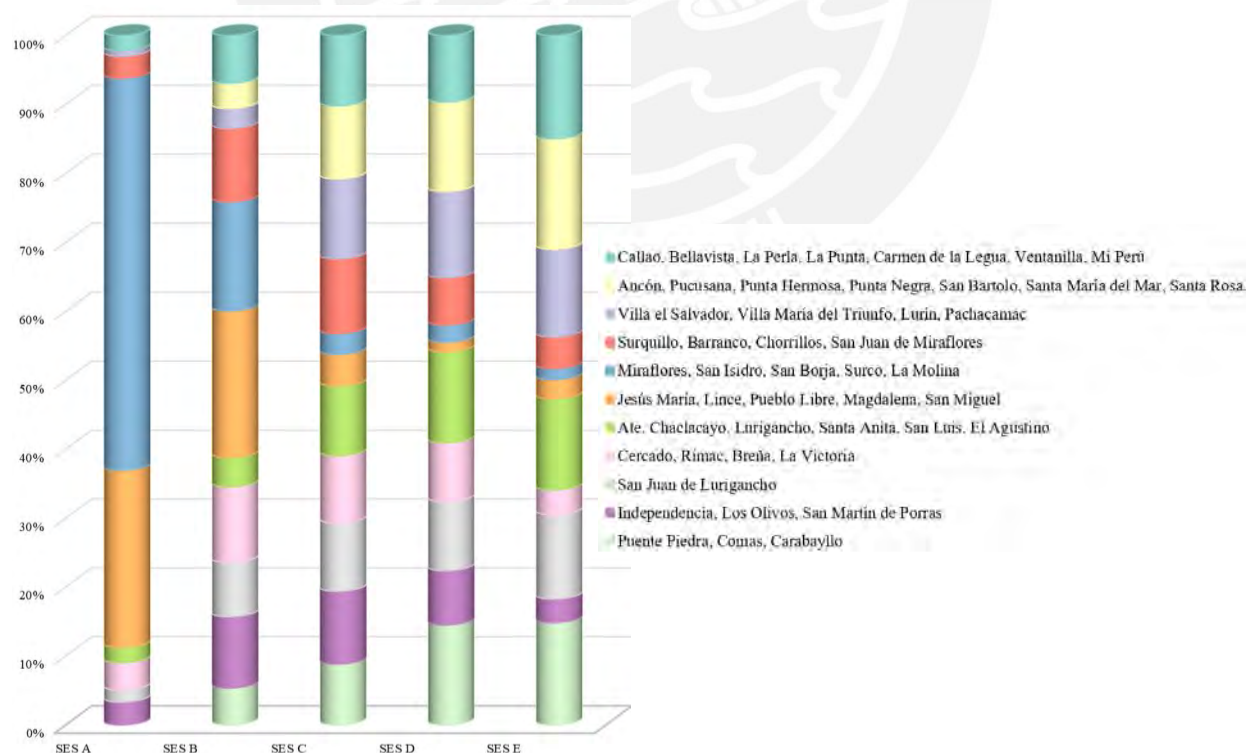


Figura 12 - Distribución de distritos por grupos según el porcentaje de nivel socioeconómico (NSE) en 2019, adaptado de (APEIM, 2019).

Por otro lado, la Figura 13 muestra la distribución de los residuos sólidos alimenticios generados en los hogares durante el año 2019, según la plataforma SIGERSOL con base en el estudio de caracterización de residuos sólidos propuesto por el Ministerio del Ambiente (MINAM, 2019a). Se aprecia que la mayor generación de residuos se encuentra presente en la periferia de la ciudad, mientras que el centro de la ciudad exhibe distritos con una menor generación.

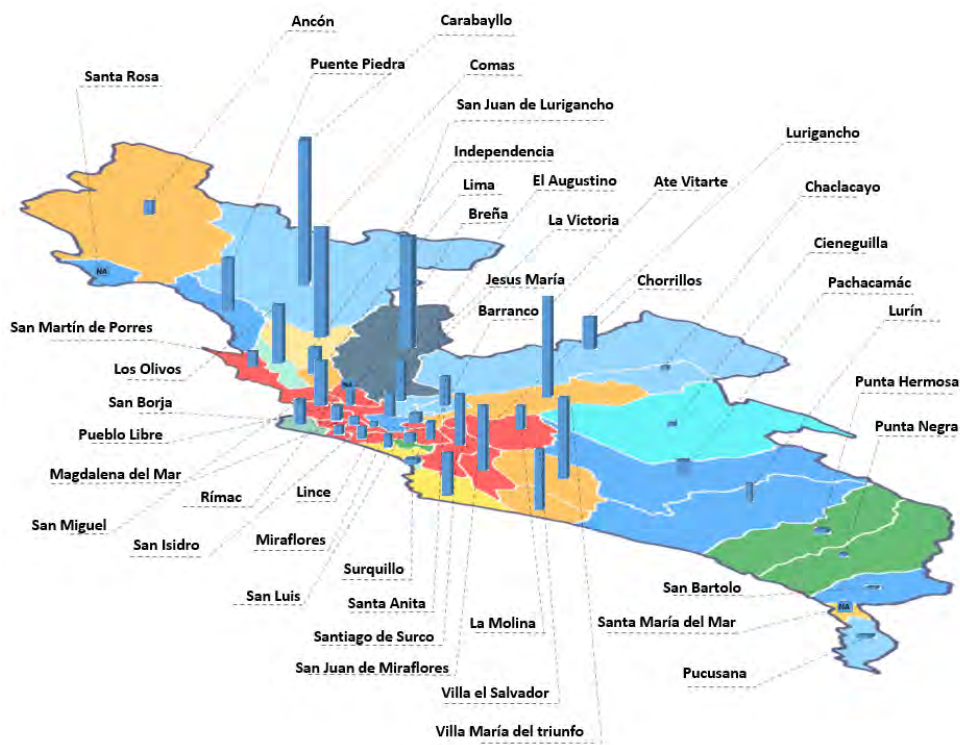


Figura 13 - Residuo de alimentos en hogares de Lima Metropolitana en 2019 (toneladas métricas / año). Datos tomados de SIGERSOL (MINAM, 2019b).

En relación con las condiciones ambientales asociadas a los residuos sólidos, especialmente en las actividades de recolección y disposición, es importante señalar que en el Perú aún persiste una gestión inadecuada de los residuos (Margallo et al., 2019). De hecho, hasta el año 2021, un total de 1970 ha de tierras permanecían degradadas por vertederos a cielo abierto en todo el país (Cristóbal et al., 2022), generando numerosos riesgos ambientales y de salud, y convirtiéndose en puntos críticos de acumulación al carecer de implementación tecnológica. En el 2017, solo existían 29 rellenos sanitarios en Perú, ubicados en 15 de las 25 regiones

(Ziegler-Rodriguez et al., 2019). Aunque, para el 2021, este número había aumentado a 65 rellenos sanitarios, ampliando la cobertura a varias ciudades de tamaño medio, la situación general del país sigue siendo delicada (MINAM, 2021). A pesar de este sombrío panorama, la situación en Lima es sustancialmente mejor, con más del 95 % de los residuos siendo dispuestos de manera formal en rellenos sanitarios en toda la ciudad (INEI, 2021). En el año de referencia de esta investigación (i.e., 2019), se identificaron dos rellenos sanitarios conocidos como "El Zapallal" (ubicado en Carabayllo, Lima) y "Portillo Grande" (situado en Lurín, Lima), ambos gestionados por Innova Ambiental, así como el relleno sanitario "Modelo del Callao" (en Ventanilla, Callao), administrado por Petramás (MINAM, 2021).

Respecto a la inspección realizada sobre el consumo de combustibles para cocinar, la Figura 14 muestra la distribución espacial de la preferencia de consumo de combustible. En ese sentido, debido al costo y la disponibilidad, el uso del GLP es predominante en toda Lima, especialmente en la actualidad (OSINERGMIN, 2020b). En contraste, el uso de GN se observa principalmente en los distritos periféricos del norte y sur. Sin embargo, considerando los esfuerzos del gobierno por expandir las redes de distribución de GN en todo el territorio peruano, especialmente en áreas residenciales, se espera que este combustible se convierta en el predominante en el futuro, superando al GLP (OSINERGMIN, 2020a).

En la actualidad, aproximadamente el 25 % de los hogares utiliza GN como fuente de combustible para cocinar. El uso de electricidad para cocinar es mínimo (i.e., 1.7 %) y se encuentra principalmente en el centro de la ciudad, especialmente en aquellos distritos con mayor poder adquisitivo; sin embargo, el uso de electricidad podría experimentar una disminución en las preferencias de consumo debido a los recientes incrementos en las tarifas eléctricas (OSINERGMIN, 2021a). Asimismo, de acuerdo a la muestra obtenida (OSINERGMIN, 2021a), de las 1797 viviendas registradas, 1201 tenían acceso al GLP, 502 tenían acceso al GN y 1793 contaban con instalación eléctrica.

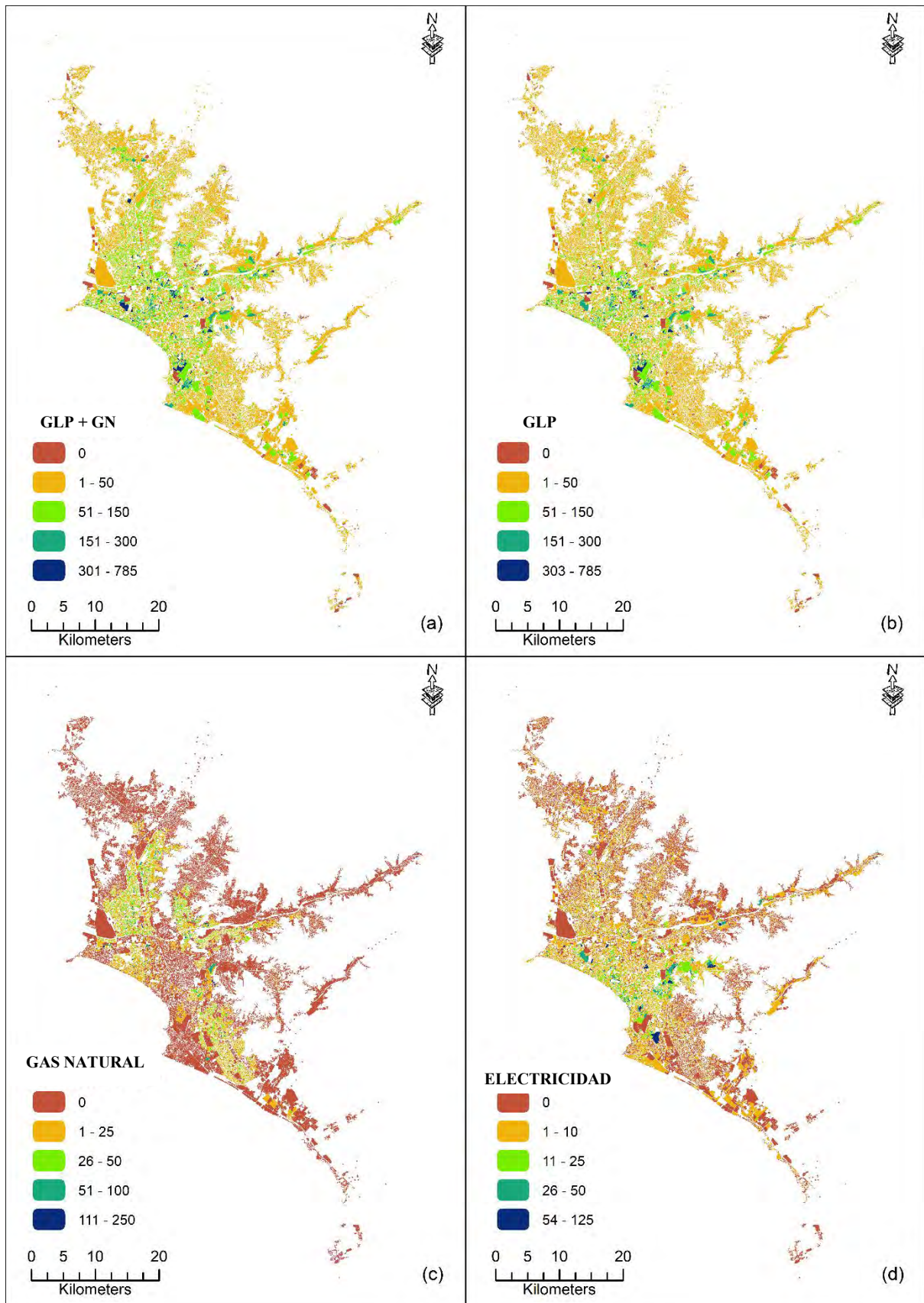


Figura 14 - Frecuencia a nivel de manzana de la fuente de energía utilizada para cocinar en la ciudad de Lima - 4a: GLP + GN, 4b: GLP, 4c: GAS NATURAL, 4d: ELECTRICIDAD. (GLP: Gas Licuado de Petróleo, GN: Gas Natural).

De la misma forma, en cuanto a las características de la población, se obtuvo que las familias están conformadas en promedio por tres personas residentes en el hogar, con un máximo de hasta diez personas. Esto implica que la densidad poblacional por vivienda también desempeña un rol importante en la generación de residuos sólidos. La Figura 15 muestra la distribución de la variable una vez hecha la limpieza de datos (i.e., retiro de valores atípicos). Para una mejor comprensión de esta variable, fue agrupada en 3 clases en función de sus percentiles, obteniendo que la clase 0 corresponde a viviendas con 1 a 2 personas, clase 1, con 2 a 4 personas por vivienda y la clase 3 con más de 4 habitantes por vivienda. Esta nueva caracterización permitirá una mejor comprensión de las variables en el contexto estudiado.

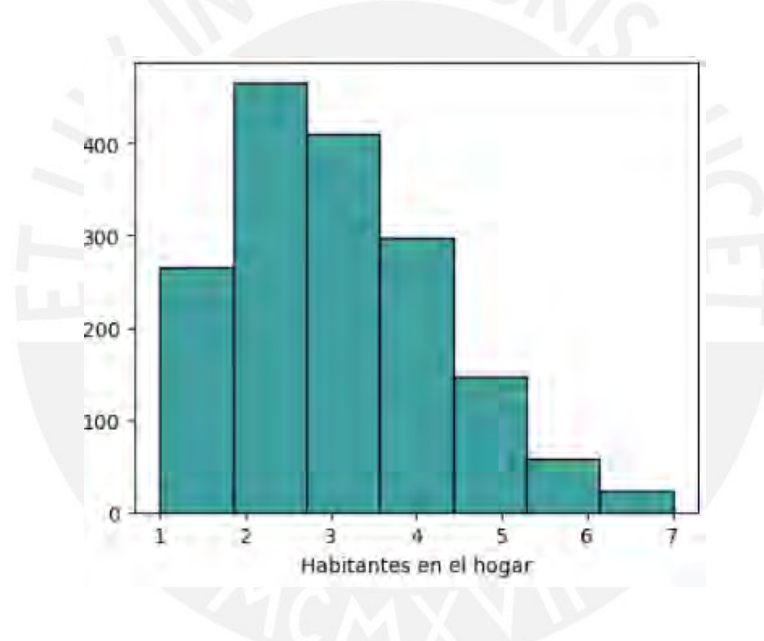


Figura 15 - Distribución de personas que habitan en la vivienda. Información elaborada a partir de ERCUE 2019.

Por otro lado, en términos de edad, este proyecto considera al jefe de familia como representante por vivienda. Bajo esta premisa, la Figura 16 presenta de manera visual el histograma correspondiente a la distribución de edades obtenida en la muestra. A partir de esta representación gráfica, se puede inferir que la distribución de edades tiende a ajustarse a una forma normal, destacando una concentración significativa alrededor de los 54 años como edad promedio.

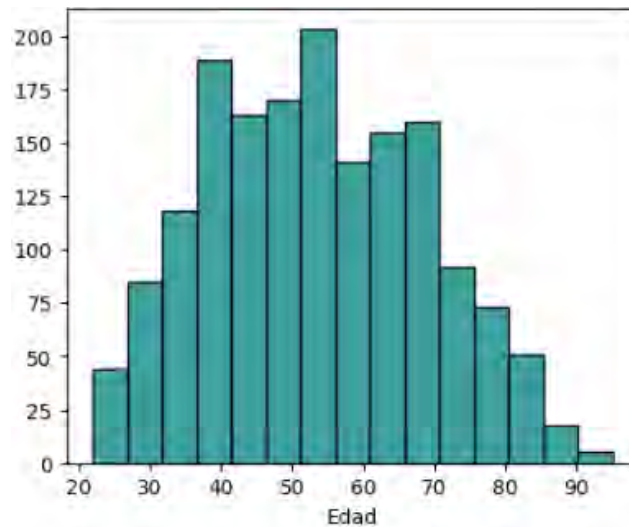


Figura 16 - Histograma de la edad de las personas consideradas como jefe de familia. Información elaborada a partir de ERCUE 2019.

En esa misma línea, se obtuvo que el nivel de educación estaba estratificado en las siguientes clases: sin nivel, educación inicial, primaria completa, primaria incompleta, secundaria completa, secundaria incompleta, superior universitaria completa, superior no universitaria completa, superior universitaria incompleta, superior no universitaria incompleta y posgrado universitario. Sin embargo, con el propósito de generar clases balanceadas, se reagruparon en tres nuevas clases en función del grado educativo máximo alcanzado o cursado. Esta nueva clasificación tiene la siguiente distribución: 0 – hasta primaria completa, 1– hasta secundaria completa y 2 – con educación superior.

En cuanto a los resultados obtenidos en la exploración de datos socioeconómicos, se destaca la notable dispersión que caracteriza la distribución de egresos (capacidad de gasto) entre los hogares analizados, lo cual se muestra en la Figura 17. Este escenario subraya la diversidad económica presente en el distrito, reflejando la variabilidad en los recursos financieros de los participantes. Además, es importante señalar la presencia de valores atípicos, aunque es importante entender que estos no necesariamente indican errores o inconsistencias en los datos, sino que, más bien, son inherentes a la naturaleza y las condiciones de vida específicas de cada distrito. En líneas generales, se aprecia un gasto promedio de 1250 soles.

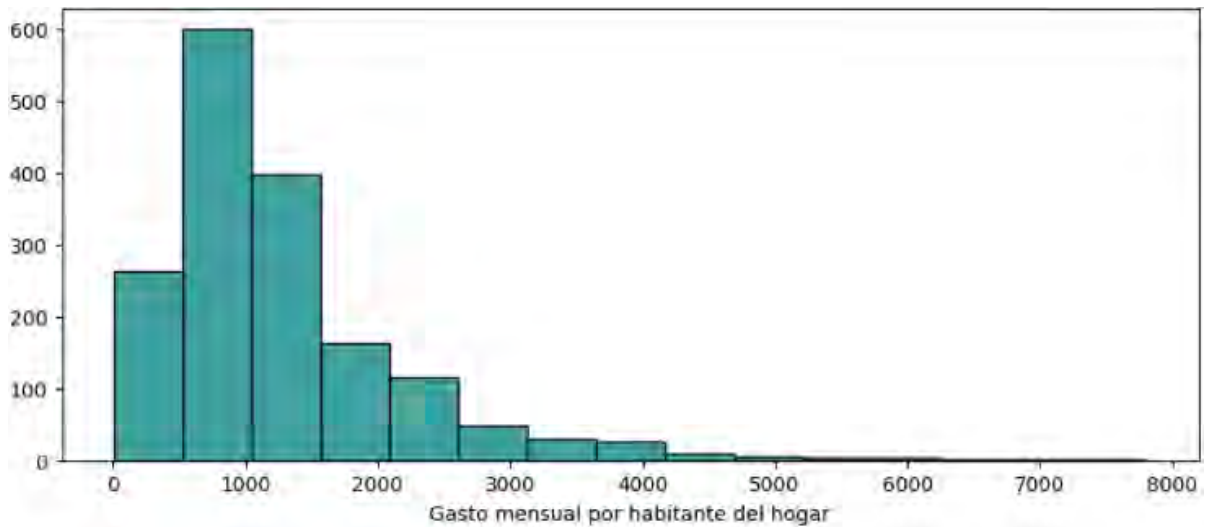


Figura 17 - Histograma del gasto mensual de las personas consideradas como jefe de familia. Información elaborada a partir de ERCUE 2019.

Finalmente, en lo que respecta a la generación per cápita de los RSM, se muestra su distribución en la Figura 18. El promedio de generación per cápita para el año de referencia se sitúa en 0.89 kg por persona por día. El histograma mostrado representa únicamente datos procesados, excluyendo valores. Además, cabe destacar que, en esta investigación, se asigna un comportamiento macro (i.e., nivel de distrito) a todas las observaciones (i.e., viviendas), según el distrito al que pertenecen.

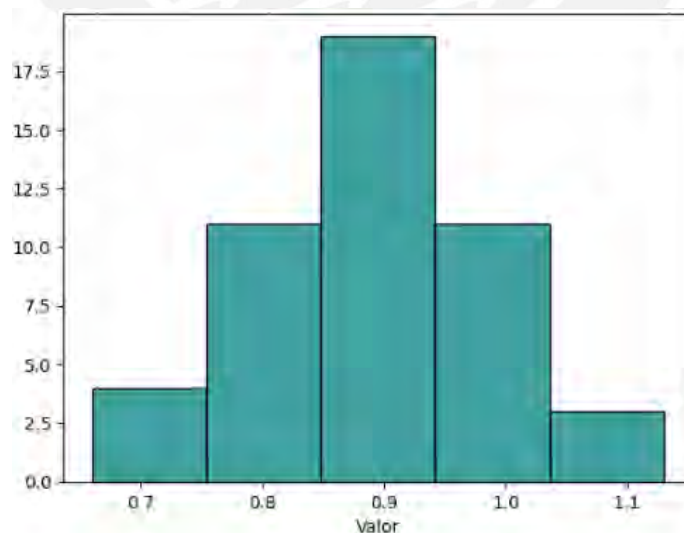


Figura 18 - Histograma referente a la generación de residuos sólidos municipales per cápita a nivel distrital de la provincia de Lima Metropolitana - 2019.

5.3 Resultados derivados del análisis exploratorio

Como resultado del análisis exploratorio llevado a cabo con base en la literatura científica y en el caso de estudio, se encontró que la generación de residuos alimentarios domésticos está influenciada por diversos factores atribuidos a la composición demográfica, características del hogar y patrones de consumo. Es importante señalar que también existen factores externos que condicionan la generación de residuos alimentarios domésticos, como los hábitos de consumo y las condiciones climáticas, ambos fuera del alcance de esta investigación.

En términos generales, el nivel socioeconómico juega un papel significativo en la generación de residuos sólidos alimentarios (i.e., a mayor nivel socioeconómico, mayor cantidad de residuos alimentarios generados). Sin embargo, considerando que el contexto económico interno de cada distrito es heterogéneo, la cantidad de residuos alimentarios generados puede ser mayor o menor que en otros distritos con características similares. La Figura 19 muestra el nivel de ajuste de cada variable identificada en esta sección frente a la generación de residuos alimentarios domiciliarios. Las variables con un mayor nivel de ajuste positivo son las asociadas al consumo de combustible conjunto de GLP y GN, esto debido a que esta variable captura las preferencias en el consumo de combustible dentro del área de estudio.

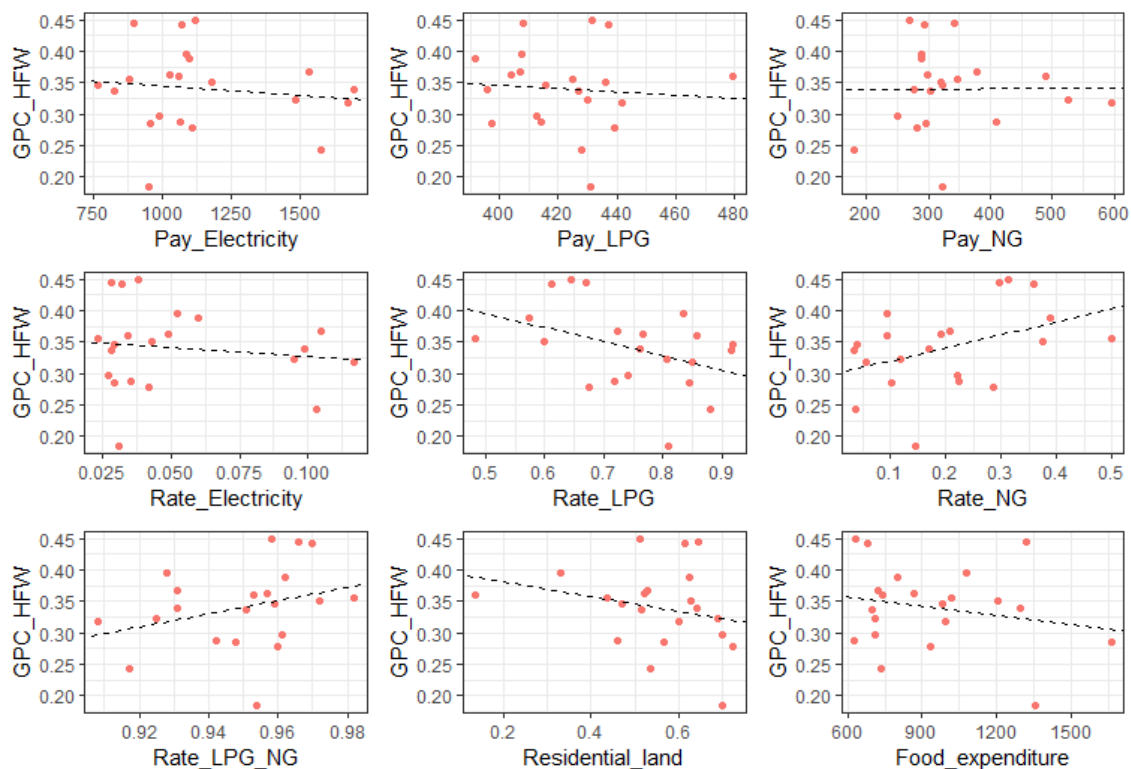


Figura 19 - Análisis de regresión del consumo de combustible para cocinar en la generación doméstica de residuos alimenticios (GPC_HFW).

Acrónimos: Residential_land: Tasa de uso de suelo residencial por distrito, Food expenditure: Gasto mensual promedio en consumo de alimentos, Rate_LPG: Tasa de uso de hogares que utilizan GLP para cocinar alimentos, Pay_LPG: Gasto mensual promedio para este combustible, Rate_Electricity: Tasa de uso de hogares que utilizan electricidad para cocinar alimentos, Pay_Electricity: Gasto mensual promedio para este combustible disponible, Rate_GN: Tasa de uso de hogares que utilizan GN para cocinar alimentos, Pay_GN: Gasto mensual promedio para este combustible, Rate_LPG_GN: Tasa de uso de hogares que utilizan GLP y GN para cocinar alimentos.

En cuanto al crecimiento demográfico y al desarrollo urbano, cuando se observa los distritos en Lima, la mayoría muestra una predominancia de áreas residenciales, especialmente en los distritos más poblados (i.e., San Juan de Lurigancho y San Martín de Porres). Sin embargo, también hay una fracción de distritos donde predominan áreas comerciales e industriales (i.e., Lurín y La Victoria). Además, la variabilidad de los NSE presentes en el estudio de caso también condiciona aspectos asociados con el gasto en alimentos y actividades económicas, reflejando la naturaleza dinámica e interrelacionada de las variables analizadas. En la misma línea, la gestión del suelo también está vinculada a los patrones de consumo, debido al desplazamiento de áreas residenciales a comerciales y/o industriales en ciertos momentos del

día y a migraciones a más largo plazo que resultan en cambios espaciales en la estructura económica de diferentes distritos, lo que se evidencia en términos de hábitos de consumo. La Figura 20 muestra el esquema relacional – causal encontrado para el caso de estudio.

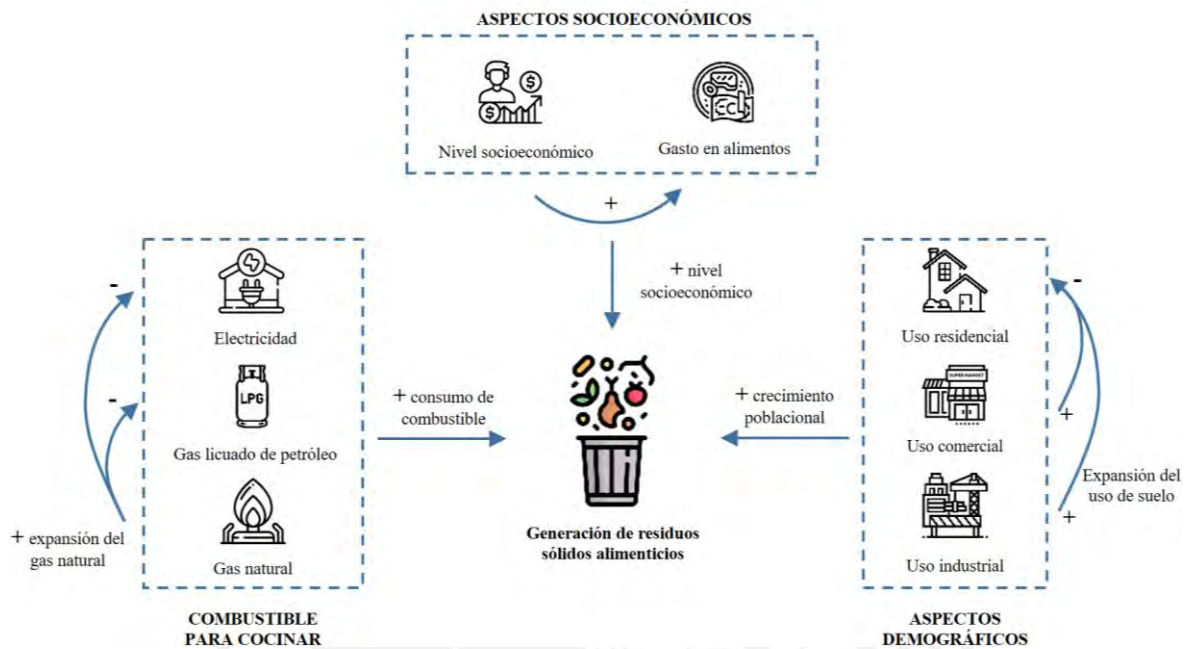


Figura 20 – Síntesis de la exploración del análisis causal de las principales variables que influyen en la generación de residuos alimentarios domésticos, basada en la evaluación realizada.

En esta representación, el símbolo "+" indica un incremento en una variable específica, mientras que el símbolo "-" indica una disminución en esa variable particular.

5.4 Selección de variables predictoras ajustadas al caso de investigación

Considerando la acelerada industrialización y rápida urbanización en Lima Metropolitana, se hace evidente la necesidad de desarrollar un sistema integral de gestión de RSM. Sin embargo, este desarrollo se ve obstaculizado por la falta o limitada disponibilidad de datos. Por lo tanto, la predicción de la generación de residuos sólidos se presenta como una oportunidad de mejora, ya que constituye una estrategia integral para abordar los problemas derivados de la deficiente gestión de los RSM debido a la calidad de los datos. En este sentido, es relevante emplear variables dinámicas, sólidas y confiables para realizar una predicción precisa de la generación de RSM, lo que a su vez contribuirá a la implementación de un sistema de gestión integrado. Estas variables desempeñarán un papel fundamental en la predicción de la generación de RSM

en nuestra área de análisis. Su elección y definición son esenciales para la precisión y pertinencia de nuestro modelo de predicción. La Tabla 5 presenta el conjunto de variables propuestas específicamente seleccionadas para el caso de estudio.

Tabla 5 - Propuesta de variables predictoras de la generación de RSM. Adaptado de Izquierdo-Horna, Kahhat, Vázquez-Rowe, (2023).

Dimensión	Variable
Sociocultural	Edad
	Nivel de educación
	Densidad de población
Económica	Gasto mensual en soles de la vivienda
	Consumo de energía/combustible mensual (electricidad, gas licuado de petróleo (GLP), gas natural (GN)) para cocinar
	Pago de energía/combustible mensual (electricidad, gas licuado de petróleo (GLP), gas natural (GN)) utilizado para cocinar en soles
Ambiental	Generación per cápita de residuos sólidos municipal

Consideramos que la sinergia entre las variables económicas, socioculturales y ambientales permitirá no solo mejorar la toma de decisiones, sino que cada dimensión en particular servirá como una herramienta para formular estrategias que fortalezcan la gestión de RSM. Además, estos resultados preliminares sirven como un punto de partida para determinar si la información puede obtenerse de bases de datos gubernamentales o si es preferible utilizar encuestas. Esta última opción conlleva importantes ventajas, como el control del proceso de captura de información y niveles bajos de incertidumbre, aunque se espera que sea ligeramente más costosa en comparación con la captura de datos tradicional (i.e., censos o encuestas nacionales).

5.5 Resultados derivados de la regresión lineal múltiple

Se empleó la regresión lineal múltiple para analizar las relaciones entre el consumo de combustible, el uso del suelo, el gasto en alimentos vinculado al consumo doméstico y la generación de residuos alimentarios domésticos, con el fin de examinar sus interconexiones.

5.5.1. Escenario 1 – Combustible predominante: GLP y electricidad

En este primer escenario, se realizó un análisis inicial para examinar la relación entre dos variables clave: el consumo de GLP y electricidad, y su impacto en la generación de residuos alimentarios domésticos. Al seleccionar estas variables, se buscó obtener datos sobre la influencia de diferentes combustibles para cocinar en los residuos alimentarios, ya que se planteó la hipótesis de que los residuos alimentarios están altamente correlacionados con el uso de combustibles para cocinar.

En la Figura 21a, se muestra la correlación existente entre las variables seleccionadas. Por un lado, las variables relacionadas con el consumo de electricidad mostraron una correlación positiva con la generación de residuos alimentarios domésticos (i.e., a mayor consumo de electricidad, mayor generación de residuos alimentarios domésticos). Con respecto al nivel de significancia de los indicadores seleccionados, se consideraron significativas las variables asociadas al consumo de GLP para cocinar con un valor de $p < 0.05$ y el uso de suelo residencial con un valor de $p < 0.10$ (Tabla 6A). Por otro lado, se encontró que las variables asociadas al consumo de electricidad no eran significativas. Por lo tanto, no contribuyen a la predicción porque están altamente correlacionadas entre sí.

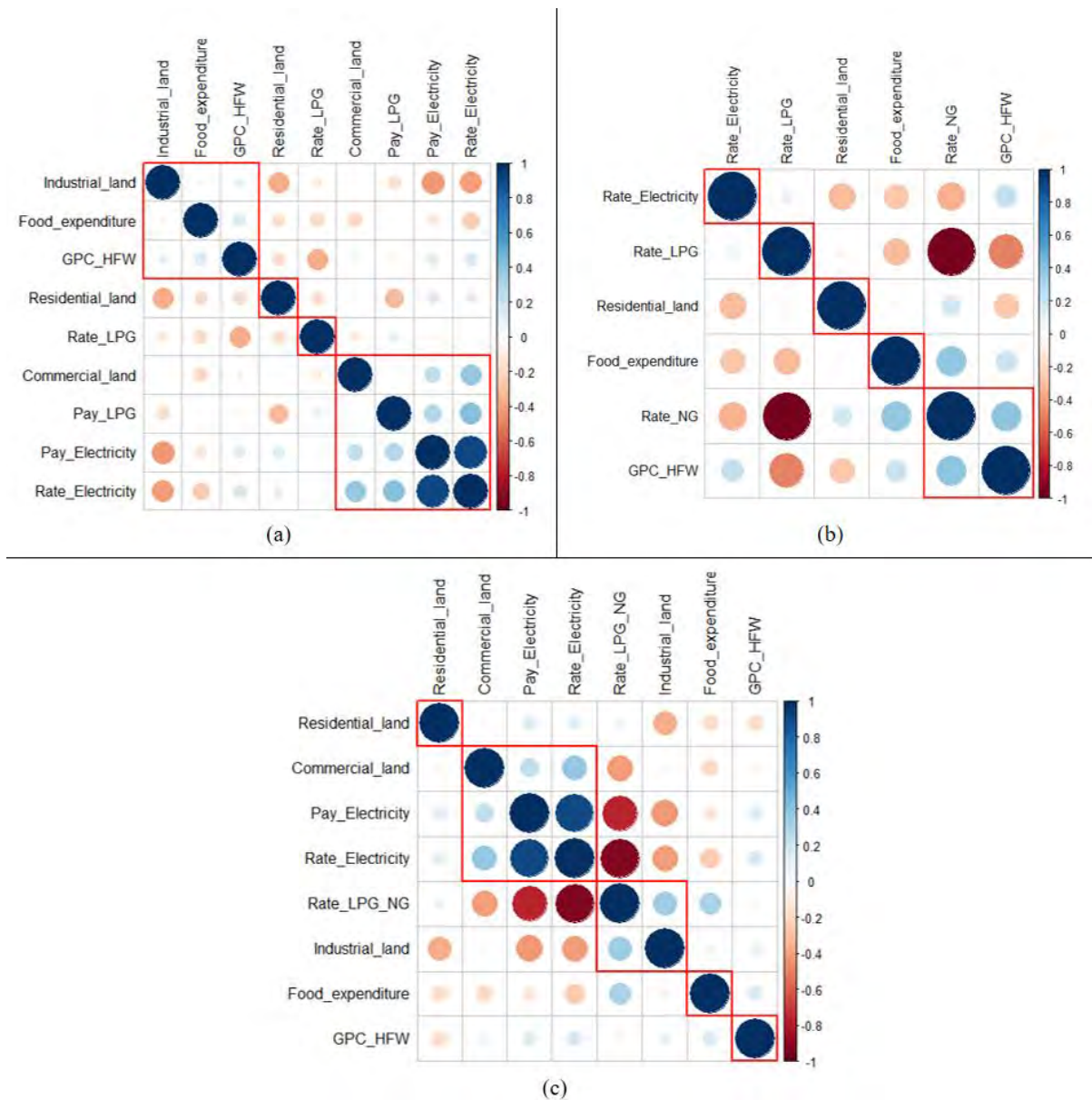


Figura 21 - Matriz de correlación de las variables utilizadas en el análisis de regresión múltiple: (a) Escenario 1: Consumo predominante: Gas licuado de petróleo y electricidad - (b) Escenario 2: Consumo de gas natural preferentemente - (c) Escenario 3: Consumo de combustible: GN + GLP considerado como un único tipo de combustible.

El tamaño de los círculos corresponde a la magnitud de la correlación entre las variables utilizadas. Los círculos más grandes indican una correlación más fuerte, siendo el color azul representativo de una correlación positiva y el color rojo mostrando una correlación negativa. Los cuadrados rojos agrupan variables altamente correlacionadas, ofreciendo una representación visual de grupos relacionados con variables estrechamente asociadas dentro de cada perspectiva analítica.

Acrónimos: Residential_land: Tasa de uso de suelo residencial por distrito, Food_expenditure: Gasto mensual promedio en consumo de alimentos (GPC_HFW), Rate_LPG: Tasa de uso de hogares que utilizan GLP para cocinar alimentos, Pay_LPG: Gasto mensual promedio para este combustible, Rate_Electricity: Tasa de uso de hogares que utilizan electricidad para cocinar alimentos, Pay_Electricity: Gasto mensual promedio para este combustible disponible, Rate_GN: Tasa de uso de hogares que utilizan GN para cocinar alimentos, Pay_GN: Gasto mensual promedio para este combustible, Rate_LPG_GN: Tasa de uso de hogares que utilizan GLP y GN para cocinar alimentos.

Los resultados obtenidos en este primer enfoque revelan una correlación entre el uso del suelo residencial, el consumo de GLP y la tasa de generación de residuos alimentarios domésticos. Además, tanto el uso del suelo como el consumo de GLP demuestran ser significativos para explicar la generación per cápita de residuos alimentarios domésticos. Aunque se esperaría una correlación positiva entre un mayor uso del suelo residencial y la generación de residuos alimentarios domésticos, debido a la densidad de población, los hallazgos del primer escenario indican una tendencia opuesta. A medida que disminuyen el uso del suelo residencial y el consumo de GLP, la tasa de generación de residuos alimentarios domésticos realmente aumenta. Este resultado inesperado se atribuye principalmente a la sustitución de GLP por GN en ciertos distritos, particularmente en áreas con una mayor proporción de uso de suelo residencial.

5.5.2. Escenario 2 – Consumo de gas natural preferentemente

El segundo escenario se centra en observaciones que implican el consumo de GN para cocinar. El conjunto de datos comprende observaciones donde el GN estaba presente al menos como combustible para cocinar. Las variables relacionadas con el consumo de GN muestran una correlación positiva con la generación de residuos alimentarios (Figura 21b), es decir, a mayor consumo de GN, mayor generación de residuos alimentarios domésticos. Estos indicios complementan el análisis causal exploratorio inicial, resaltando la interrelación entre estas variables y la influencia significativa del nivel socioeconómico en la preferencia por un combustible específico para cocinar. Por ejemplo, la asequibilidad y conveniencia del GN como fuente de energía para cocinar pueden llevar a su mayor adopción en los hogares. Finalmente, los resultados del modelo optimizado muestran que los indicadores asociados al consumo de GN son los más relevantes con un valor de $p < 0.05$ (Tabla 6B).

Tabla 6 - Resultados del análisis de regresión múltiple para la generación per cápita de residuos alimentarios en los hogares. Adaptado de Izquierdo-Horna, Kahhat, Vázquez-Rowe, (2024)

(A) - Resultados del análisis de regresión para GPC_HFW (Generación per cápita de residuos de alimentos en los hogares) considerando sólo las variables asociadas al consumo de GLP y electricidad.					
Residuals	Min	IQ	Median	3Q	Max
	-0.121	-0.033	-0.006	0.034	0.120
	Coefficients	Std. error	t-value	Pr (> t)	
Intercept	0.676	0.142	4.761	2.8 E-05	
Pay_Electricity	-0.125	5.1 E-05	-0.245	0.807	
Pay_LPG	-2.9 E-04	2.4 E-04	-1.195	0.240	
Rate_Electricity	0.372	3.7 E-01	1.006	0.321	
Rate_LPG	-0.208	7.3 E-02	-2.851	0.007	
Food_expenditure	-1.3 E-05	2.9 E-05	-0.430	0.669	
Residential_land	-8.7 E-02	4.6 E-02	-1.883	0.067	
(B) - Resultados del análisis de regresión para GPC_HFW (generación per cápita de residuos de alimentos en los hogares) considerando sólo las observaciones que al menos utilizan GN - GN: Gas Natural					
Residuals	Min	IQ	Median	3Q	Max
	-0.085	-0.045	0.002	0.029	0.080
	Coefficients	Std. error	t-value	Pr (> t)	
Intercept	0.311	0.022	14.32	2.7 E-11	
Rate_NG	0.182	0.089	2.055	0.050	
(C) - Resultados del análisis de regresión múltiple para GPC_HFW (generación de residuos de alimentos domésticos per cápita) considerando GN+GLP como uno - GLP: gas licuado de petróleo, GN: gas natural.					
Residuals	Min	IQ	Median	3Q	Max
	-0.125	-0.031	-0.010	0.021	0.123
	Coefficients	Std. error	t-value	Pr (> t)	
Intercept	-0.585	0.601	-0.972	0.337	
Rate_Electricity	0.850	0.460	1.849	0.072	
Rate_LGP_GN	0.961	0.622	1.545	0.130	
Residential_land	-0.069	0.048	-1.441	0.157	

Acronimos: Residential_land: Tasa de uso de suelo residencial por distrito, Food expenditure: Gasto mensual promedio en consumo de alimentos, Rate_LPG: Tasa de uso de hogares que utilizan GLP para cocinar alimentos, Pay_LPG: Gasto mensual promedio para este combustible, Rate_Electricity: Tasa de uso de hogares que utilizan electricidad para cocinar alimentos, Pay_Electricity: Gasto mensual promedio para este combustible disponible, Rate_GN: Tasa de uso de hogares que utilizan GN para cocinar alimentos, Pay_GN: Gasto mensual promedio para este combustible, Rate_LPG_GN: Tasa de uso de hogares que utilizan GLP y gas natural para cocinar alimentos.

5.5.3. Escenario 3 – Combustible predominante: GLP y GN considerado como un único tipo de combustible

El tercer escenario implica trabajar con una variable construida que considera la suma del consumo de GN con GLP. La agregación de ambos parámetros muestra una correlación positiva con los residuos alimentarios domésticos (Figura 21c). Además, la variable artificial

toma una relevancia significativa con un valor de $p < 0.15$. Es importante tener en cuenta que el valor de p solo mide la evidencia estadística y no ofrece perspectivas vinculadas a la magnitud o importancia práctica de la relación entre variables (Andrade, 2019). Con respecto al nivel de significancia de los indicadores seleccionados, las variables asociadas al uso de electricidad también muestran una relación positiva con la generación de residuos con un valor de $p < 0.10$ (Tabla 6C).

En suma, se puede inferir que un aumento en el consumo de combustible (i.e., electricidad, GN, GLP) conduce a una mayor generación de residuos alimentarios domésticos. Esta correlación se respalda aún más por el hecho de que la facilidad de cocinar y la disponibilidad constante de combustible podrían fomentar un mayor consumo de alimentos procesados y precocidos. Sin embargo, más allá de los residuos orgánicos, cabe destacar que este tipo de elecciones alimentarias a menudo también generan envases desechables y residuos adicionales (Kan & Miller, 2022).

5.5.4. Consideraciones adicionales

Basándonos en los niveles de significancia observados en el análisis de regresión y las relaciones identificadas en el análisis exploratorio, planteamos la hipótesis de que la generación de residuos alimentarios domésticos está principalmente vinculada con los siguientes parámetros: poder adquisitivo, consumo de combustibles para cocinar (i.e., GN + GLP y electricidad) y uso del suelo. Además, la participación y conciencia ciudadana en términos de responsabilidad ambiental también deben ser consideradas (Hoehn et al., 2023) para comprender la generación de residuos alimentarios domésticos y desarrollar estrategias de conversión de residuos en recursos.

Por otro lado, para contribuir de manera efectiva a las herramientas de cuantificación e identificación de residuos alimentarios domésticos, debemos tener en cuenta las motivaciones

y limitaciones de la población. Esto incluye factores como el conocimiento sobre el cuidado del ambiente y la disponibilidad de infraestructura (Graham-Rowe et al., 2014). Además, comprender también la relación entre los residuos alimentarios domésticos y la ocupación del suelo es de gran importancia. La utilización de tierras con fines residenciales es un factor clave para idear estrategias de gestión de residuos, ya que requiere una comprensión profunda de los impulsores del comportamiento alimentario doméstico (Ananda et al., 2023). En este sentido, comprender cómo el uso del suelo y la densidad de población impactan en la generación de residuos alimentarios domésticos permite identificar oportunidades de reducción de residuos, implementar intervenciones y abogar por políticas sostenibles de planificación urbana que prioricen la infraestructura de gestión de residuos y prácticas respetuosas con el ambiente (D. Y. Xu et al., 2016).

5.6 Resultados derivados del modelo predictivo implementado: *random forest*

En esta sección, se lleva a cabo la evaluación de las métricas de validación, explorando la importancia de las variables predictoras y posibles patrones de error. Además, se presentan visualizaciones adicionales de los datos utilizados.

5.6.1. *Análisis complementario de la información recopilada y balanceo de clases*

Como parte de un análisis integral basado en la implementación del *random forest*, esta sección se centra en profundizar la interpretación y comprensión de la información recopilada, consolidando así los resultados obtenidos mediante la aplicación del algoritmo mencionado. Este análisis complementario busca proporcionar una visión más detallada de las sinergias de las variables de entrada, contribuyendo de manera significativa a la robustez y claridad de las conclusiones derivadas de la investigación. De la misma forma, considerando la naturaleza de los datos y el concepto de clases balanceadas —es decir, categorías que difieren considerablemente en el número de elementos— se planteó una nueva categorización para determinadas variables en función de sus percentiles.

La Tabla 7 muestra la totalidad de variables trabajadas con este fin. Asimismo, es importante mencionar que existen técnicas recomendadas para el balanceo de clases como las que cambian la distribución de los datos con la finalidad de balancear las categorías y las que tratan de variar los algoritmos para mejorar la predicción. El primer grupo de técnicas abarca el submuestreo de las clases mayoritarias o el sobremuestreo de las clases minoritarias (Véliz, 2023). Si bien es cierto que estas técnicas suelen ayudar a compensar clases minoritarias, traen consigo pérdida de información o alteración de la base de datos original; por ello, en este proyecto solo se está empleando la categorización alternativa de las variables exclusivamente para fines de visualización e identificación de sinergias entre las variables utilizadas para la implementación del algoritmo de *random forest*.

Tabla 7 - Variables tratadas por presentar clases no balanceadas. Se reagruparon en nuevas categorías en función de los percentiles de cada una.

Variable	Clases originales	Clases generadas
Número de habitantes en el hogar (<i>Number_of_HM</i>)	mean 3.037	
	std 1.540	
	min 1.000	0: Baja concentración
	25 % 2.000	1: Media concentración
	50 % 3.000	2: Alta concentración
	75 % 4.000	
Información a nivel vivienda	max 10.00	
	Secundaria completa - 4216	0: Nivel de educación baja
	Primaria completa - 1987	1: Nivel de educación medio
	Primaria incompleta - 1873	2: Nivel de educación alto
	Secundaria incompleta - 1686	La categoría 0 contempla a primaria completa como nivel máximo alcanzado.
	Superior universitaria completa - 1497	La categoría 1 contempla a secundaria completa como nivel máximo alcanzado.
	Superior no universitaria completa - 1226	La categoría 2 contempla a post grado universitario como nivel máximo alcanzado.
	Sin nivel - 817	
	Superior universitaria incompleta - 394	
	Superior no universitaria incompleta - 377	
Posgrado universitario - 99		
Educación inicial - 28		

La Figura 22 muestra el histograma que se obtiene al relacionar la generación de RSM por vivienda (*RSM_VIVIENDA*) segmentada por el número de habitantes del hogar (*Number_of_HM*). Podemos apreciar que la mayor generación de residuos sólidos se encuentra vinculada a aquellas viviendas con un mayor número de habitantes en el hogar, lo cual es coherente con lo revisado en la literatura. De la misma manera, viviendas con menor generación de residuos se encuentran asociados a un menor número de residentes.

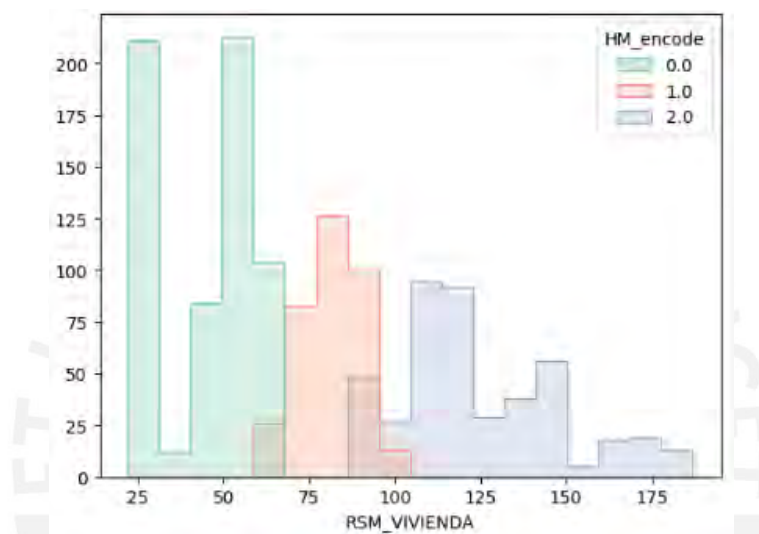


Figura 22 - Histograma que representa la distribución de la variable *RSM_VIVIENDA* agrupadas según la categoría "*HM_encode*". (*HM_encode*: Clases generadas en función del número de habitantes del hogar)

La Figura 23 muestra el histograma que se obtiene al relacionar la generación de RSM por vivienda (*RSM_VIVIENDA*), gasto mensual de la persona considerada como jefe de familia (*Expense_pers*) segmentada por el nivel de educación. La exploración de la capacidad de gasto en las familias revela un panorama dinámico en el contexto del estudio. Es evidente que aquellas que se sitúan en el rango de 0 a 2000 soles exhiben una notoria variabilidad, reflejo directo del vibrante dinamismo económico característico del caso de estudio. Este fenómeno sugiere una diversidad significativa en los hábitos de gasto de estas familias, posiblemente influenciada por factores socioeconómicos y culturales. A medida que ascendemos en el espectro de capacidad de gasto, desde 4000 soles en adelante, se observa una disminución en la variabilidad. Este patrón podría asociarse a niveles educativos más elevados, corroborando

la literatura que sugiere una correlación entre educación superior y una menor generación de RSM.

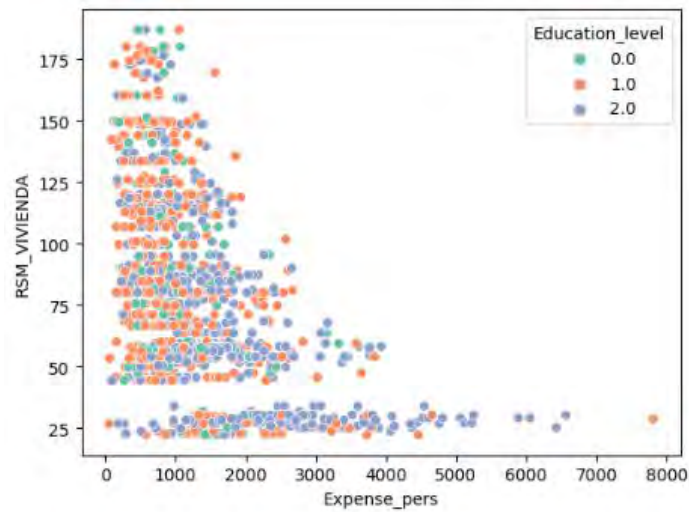


Figura 23 - Histograma que representa la distribución de la generación RSM por vivienda (*RSM_VIVIENDA*) y el gasto mensual por persona (*Expense_pers*) agrupadas según la categoría "*Education_level*" (*Education_level*: Clases generadas en función del nivel de educación modal presente en la vivienda)

5.6.2. Optimización del modelo: Hiperparámetros

La optimización de los hiperparámetros desempeña un papel esencial en la mejora de la capacidad predictiva y la generalización del modelo, garantizando su adecuación a las características específicas de los datos analizados. Para el modelo implementado, se procedió a determinar la cantidad óptima de árboles, la profundidad de estos y el criterio de división de cada nodo. Como resultado de la búsqueda grillada, se observa que se alcanza una estabilización del modelo al utilizar 150 árboles. Asimismo, para la configuración obtenida se dejó sin restricciones la profundidad máxima de los árboles, permitiendo que los nodos se expandan hasta que contuvieran un número mínimo de muestras o hasta que todas las hojas contuvieran un número mínimo de muestras requeridas. En cuanto al proceso de división de nodos requerirá un mínimo de dos muestras, garantizando así la construcción de un modelo robusto sin comprometer la generalización.

5.6.3. Métricas de evaluación del modelo

El rendimiento del modelo se evaluó a través del coeficiente de determinación (R^2) y el error cuadrático medio (MSE). Los resultados indican que el modelo logró un R^2 de 55.6 % en la fase de entrenamiento; mientras que en la fase de testeó alcanzó un 51 %, lo que sugiere que aproximadamente el 51 % de la variabilidad en la variable objetivo puede ser explicada por las características predictoras incluidas en el modelo. Además, el MSE obtenido fue de 0.86 en la fase de testeó, lo que indica que el modelo tiene una precisión razonable en la predicción de los valores de la variable objetivo, con valores de error relativamente bajos en promedio. A continuación, se presenta un gráfico Q-Q (Quantile-Quantile) plot (Figura 24) que muestra la comparación entre las distribuciones de los valores predichos y los valores observados, lo que ayuda a evaluar la adecuación del modelo y la normalidad de los residuos. El eje X representa los cuantiles teóricos esperados y el eje Y representa los cuantiles observados de la muestra que se está evaluando (i.e., los residuos del modelo).

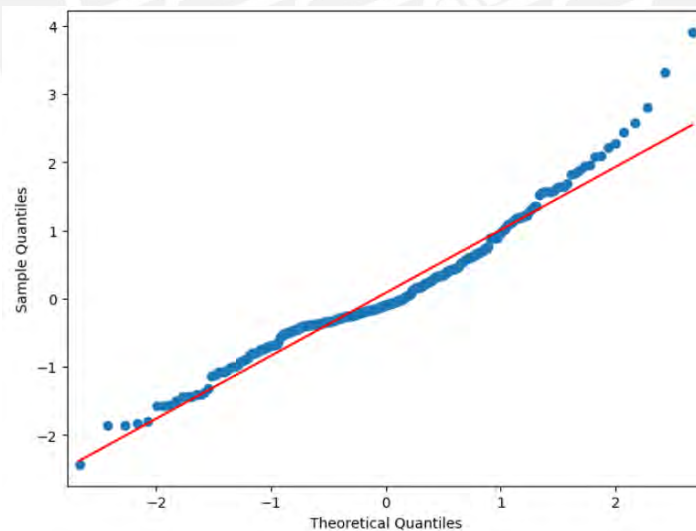


Figura 24 - Gráfico Q-Q plot que compara las distribuciones de los valores predichos y observados

5.6.4. Importancia de las variables

Se procedió a analizar la importancia relativa de las variables predictoras en el modelo *random forest*. La identificación y discusión de las variables más influyentes permiten comprender

mejor los factores que impulsan las predicciones del modelo. En este contexto, se exploraron visualizaciones que destacan la contribución de cada variable, proporcionando una visión más detallada de la relevancia de los predictores en el contexto del problema abordado. La Figura 25 muestra la importancia relativa de las variables utilizadas en el modelo de predicción. Destacan principalmente el gasto mensual de la persona, edad y pagos de combustibles.

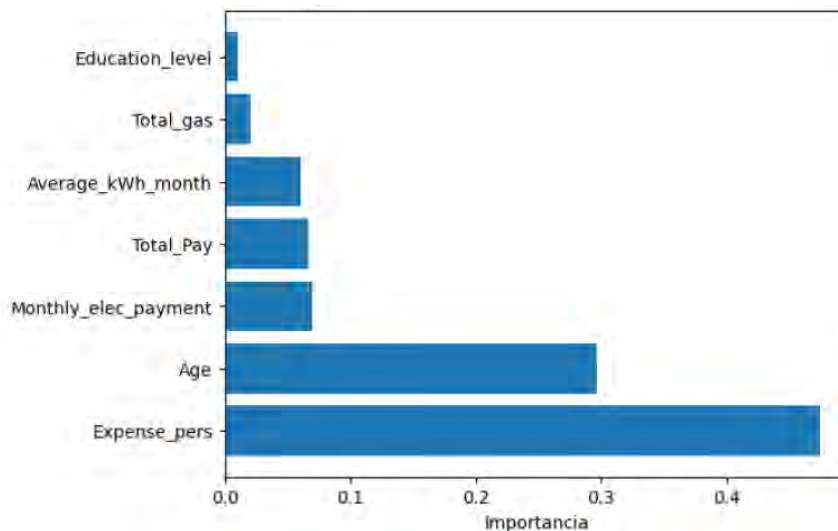


Figura 25 - Importancia relativa de las variables predictoras utilizadas en el modelo de random forest para la predicción de residuos sólidos municipales.

Acrónimos: Education_level: Nivel de educación modal del hogar, Total_gas: Consumo mensual de combustible en el hogar que unifica GLP y GN, Monthly_elec_Payment: Pago mensual por el consumo de electricidad, Average_kWh_month: Consumo mensual promedio de electricidad en el hogar, Total_pay: Pago mensual de combustible en el hogar que unifica GLP y GN, Age: Edad promedio del hogar, Expense_pers: Gasto mensual representativo por persona

5.7 Análisis de las limitaciones e incertidumbres encontradas en la ejecución metodológica

Para abordar de manera holística las limitaciones y fuentes de incertidumbre identificadas, resulta necesario reconocer la singularidad inherente a cada una de ellas y su consiguiente influencia en las prácticas efectivas de gestión de residuos. La primera subsección de este apartado se dedicará al análisis de las limitaciones en términos de orden espacial, temporal y metodológico. La segunda subsección, por su parte, se enfocará en las incertidumbres de índole epistémica identificadas durante el proceso de recolección de información.

5.7.1. Análisis de limitaciones

El proyecto de investigación, al seleccionar Lima Metropolitana como caso de estudio, se encuentra geográficamente vinculado a este territorio, lo que confiere una contextualización específica a los resultados derivados de los análisis realizados. Es necesario señalar que los resultados obtenidos son aplicables de manera exclusiva al ámbito de estudio; no obstante, la metodología empleada exhibe una reproducibilidad intrínseca, pudiendo ser extrapolada con validez a otros escenarios análogos. Asimismo, la temporalidad surge como un componente de significativa relevancia, considerando la dinámica inherente a las condiciones analizadas, las cuales experimentan variaciones relacionadas a los cambios en las condiciones externas (i.e., pandemias, conflictos sociales, etc.). En este contexto, es importante destacar que la temporalidad —en este proyecto específicamente asociada al año 2019— proporciona un retrato representativo de la situación en términos de las condiciones y modos de vida de los residentes en ese periodo específico.

Por otro lado, en términos metodológicos, una limitación significativa que surge en este estudio radica en la insuficiencia de fuentes alternativas de información, lo que a menudo conduce a que la mayoría de los estudios recurran a parámetros similares sin considerar las particularidades de los casos de estudio. Este escenario complica la determinación de un conjunto de variables adecuadas para una comunidad específica, ya que, en muchos casos, las variables se han incorporado sin una validación integral de su relevancia. Como resultado, muchas de las recomendaciones se sustentan en la percepción de los autores y en la literatura previamente revisada.

Aunado a ello, resalta también la calidad y acceso de los datos. La escasa información relacionada a la generación per cápita de residuos sólidos y su composición en algunos municipios incluidos en el estudio se presenta como un apartado a tener en cuenta, ya que el desconocimiento de estos valores conduce a perder información. A pesar de que estos

municipios suelen ser menos poblados y tienen un impacto limitado en los resultados generales, esta brecha en la información obstaculiza la comprensión completa y la gestión efectiva de los desafíos de residuos en esas áreas. Además, la baja frecuencia de la recopilación de datos (que se limita a registros anuales) dificulta la supervisión continua y la toma de decisiones basadas en datos. La falta de desagregación geográfica de los datos, actualmente limitada al nivel del distrito, presenta restricciones en términos de precisión y análisis detallado. Estas limitaciones resaltan la necesidad de superar desafíos en la calidad y disponibilidad de los datos, especialmente en áreas menos pobladas, para mejorar la gestión integral de residuos sólidos y la toma de decisiones informadas.

5.7.2. *Análisis de incertidumbres*

Además de las limitaciones mencionadas, es fundamental reconocer la incertidumbre inherente a la construcción de una base de datos consolidada a partir de diversas fuentes (i.e., ERCUE & SIGERSOL). La combinación de los informes municipales con las bases de datos gubernamentales introduce cierto grado de variabilidad y posibilidad de errores, lo que impacta en la calidad de los datos resultantes. Asimismo, al considerar el comportamiento macro a nivel distrital como equivalente o promedio para todas las viviendas analizadas, se asume una homogeneidad que puede no reflejar con precisión las variaciones en la generación de residuos a nivel individual. Esta simplificación puede llevar a una cierta pérdida de detalles y subestimar la variabilidad real en la producción de RSM a nivel de viviendas.

Por otro lado, la incertidumbre epistémica vinculada al desconocimiento de los procedimientos utilizados en los estudios de caracterización de residuos sólidos o en la realización de encuestas, cuyos resultados han nutrido esta investigación, constituye un elemento fundamental en la evaluación de la fiabilidad de los datos empleados. La falta de claridad sobre la metodología y los protocolos seguidos en dichos estudios puede introducir sesgos o limitaciones en la interpretación de los resultados. Este vacío epistémico resalta la importancia de una revisión

rigurosa de la calidad y la integridad de los datos subyacentes, así como la necesidad de considerar cuidadosamente las posibles implicaciones de las decisiones metodológicas tomadas en los estudios previos.

5.8 Comparativa de los resultados obtenidos con estudios similares en la literatura

En función de los resultados obtenidos previamente, esta sección tiene como objetivo contrastar y analizar de manera sistemática las variables, técnicas, fuentes de información y resultados obtenidos en este estudio con los hallazgos de investigaciones anteriores. Este análisis comparativo busca identificar similitudes, divergencias y posibles mejoras en las metodologías y resultados, contribuyendo así a la ampliación del conocimiento en el campo de la gestión de RSM. Se explorará detalladamente cada aspecto relevante, permitiendo una evaluación crítica y fundamentada que respalde la interpretación de los resultados obtenidos en el presente estudio en relación con la literatura científica precedente.

Respecto a la implementación del modelo, Guo et al. (2021) realiza un análisis completo de la ejecución de diversos modelos de aprendizaje automático, abordando campos de aplicación específicos, características y limitaciones. Aunque destaca la popularidad de las redes neuronales, el estudio ofrece una visión integral que considera diferentes modelos en función de su idoneidad para distintos contextos y aplicaciones. Por otro lado, sugiere también que la aplicación de *random forest* es recomendable en escenarios con datos perdidos, debido a su fácil interoperabilidad. Además, también se destacan las diversas aplicaciones del aprendizaje automático, en especial, aquellas asociadas a la identificación de puntos de colección de basura (Viktorin et al., 2023).

Asimismo, de la literatura revisada, el algoritmo de red neuronal artificial es la técnica estadística más aplicada para la predicción de RSM en distintas escalas temporales (i.e., semanal, mensual, anual) (A. Xu et al., 2021). Esto es debido a que las redes neuronales exhiben

una notable capacidad de autoaprendizaje que permite mapear complejas relaciones no lineales, siendo idóneas para modelar cuestiones vinculadas a los residuos sólidos que involucran diversas fuentes de residuos, composiciones heterogéneas e impactos ambientales, económicos y sociales integrados. Sin embargo, para un buen desempeño, necesitan de una gran cantidad de datos para el entrenamiento. Además, su implementación es computacionalmente costosa y está fuertemente condicionada por la calidad y cantidad de datos disponibles, lo que complica la interpretación de las relaciones entre los parámetros de entrada y salida. En ese sentido, considerando que los datos disponibles son limitados y poseen una elevada incertidumbre epistémica, el *random forest* se erige como el modelo a elegir.

Respecto al desempeño del modelo, de acuerdo con Kannangara et al. (2018), parámetros asociados a condiciones demográficas y socioeconómicas son los recomendables para una predicción de RSM a nivel de municipios (distritos). En términos de muestra, contó con datos de 220 municipalidades. En términos metodológicos, implementaron dos modelos de aprendizaje automático: redes neuronales y árboles de decisión, concluyendo que los algoritmos de aprendizaje automático pueden ser empleados con éxito para generar modelos de residuos con un buen rendimiento de predicción. Cabe señalar que el modelo de redes neuronales alcanzó un coeficiente de determinación del 72 %, mientras que el modelo de árbol de decisión obtuvo un 54 %. En comparación, el modelo de predicción implementado mediante *random forest* mostró un R^2 del 51 %. Aunque este último valor puede parecer bajo, es necesario considerar la cantidad limitada de información disponible, representada únicamente por 50 municipalidades. Asimismo, también nos aunamos a las recomendaciones propuestas relacionadas a la generación de herramientas que contribuyan con la planificación regional de residuos mediante el abastecimiento, preprocesamiento, integración y modelado de bases de datos disponibles.

Por otro lado, respecto a las recomendaciones de utilizar el consumo de combustible como indicador relevante a tomarse en cuenta en tareas de predicción de RSM, Velusamy et al. (2023) identificó que el consumo de combustibles fósiles desempeña un rol preponderante no solo para la minimización de RSM, sino también, para el cuidado ambiental. En esa misma línea, la analogía asociada a que una mayor generación de residuo domiciliario redundaría en mayor generación municipal es soportada por Lakhout et al. (2023). Su investigación centra su fuente de información en datos de residuos domésticos y en comportamiento de generación de residuos en el hogar. Al contar con datos recopilados por una década, plantearon diversos métodos de predicción, obteniendo un ajuste de entre 67 % - 85 %. Asimismo, Abdella Ahmed et al. (2022) resalta la influencia de los niveles socioeconómicos en la generación de RSM, reforzando los hallazgos obtenidos en el análisis exploratorio de esta investigación.

5.9 Replicabilidad de la propuesta metodológica implementada

La propuesta metodológica presentada en este proyecto de investigación puede ser replicable a cualquier nivel de medición. Esto se debe a que en las suposiciones en las que se sustenta, solo son válidas si se evidencian en el área de interés. En ese sentido, la Figura 26 muestra el diagrama de flujo propuesto para una adecuada contextualización del caso de estudio con fines de predicción de RSM. Aquí sugerimos que, para poder estandarizar las metodologías de identificación de variables, es importante caracterizar la zona de estudio en dimensiones socioculturales, económicas y ambientales, esto con la finalidad de describir el comportamiento del público objetivo. Luego de ello, empezar a conocer el área de estudio desde la teoría; es decir, revisar guías, reportes, informes municipales, etc., para luego complementar el análisis contextual con visitas a la zona. Con esta primera exploración, se debe estar en la posibilidad de elaborar una lista preliminar de variables influyentes que luego, en función de la disponibilidad de datos y alcance del proyecto, servirán como punto de partida para la elaboración de suposiciones, generalizaciones y procesamiento de datos.

En relación con las variables influyentes, sugerimos que permitan segmentar la muestra. Por ejemplo, en la parte sociocultural, es importante que la población se represente en términos de edad, condiciones demográficas (i.e., densidad de población), nivel educativo o características culturales. La razón de esto está vinculada a que cada persona tiene una representación cultural específica. El aspecto ambiental puede representarse, por ejemplo, a través de variables como la humedad, la precipitación y la temperatura, ya que estas afectan los patrones de consumo y estos, a su vez, afectan la generación de RSM. Finalmente, el aspecto económico complementa adecuadamente la representación del estudio de caso, en el cual, dependiendo del objeto de análisis, la variable relacionada con el ingreso económico es la más relevante. Con las variables detalladas, se procede a la identificación de las fuentes de información para un segundo análisis conducente a la elección del método de predicción. Para ello, hay que tener en cuenta tipo de datos y número de observaciones. Finalmente, con el modelo elegido, se procede a implementar el modelo de predicción y a evaluar su desempeño.



Figura 26 – Propuesta de flujo de procesos para la implementación de metodologías estandarizadas para identificar variables predictivas de generación de RSM

CAPÍTULO 6: PERSPECTIVAS Y CONCLUSIONES

La necesidad de metodologías estandarizadas para la cuantificación y caracterización de RSM para su posterior predicción está justificada, ya que la identificación adecuada de variables influyentes en la generación de RSM es un paso fundamental y complejo que se aborda de distintas perspectivas. Además, al explorar las posibles razones del por qué no existe una estandarización de criterios, identificamos distintas limitaciones asociadas a la disponibilidad de información, confiabilidad de datos e interpretación de estos conceptos. Esta exploración inicial fue necesaria para proponer un set de variables potencialmente capaces de predecir adecuadamente la generación de RSM, especialmente en entornos urbanos. La operacionalización de estas variables aunadas a la identificación de un adecuado modelo predictivo requirió de una evaluación integral de la literatura disponible y metodologías predictivas afines.

Por otro lado, utilizando información provista por organizaciones gubernamentales, como OSINERGMIN, hemos demostrado que es factible predecir de manera indirecta la generación de RSM. Sin embargo, es esencial destacar que dicha información no es recopilada originalmente con el propósito de analizar la generación de residuos, sino que tiene un enfoque distinto. Por ejemplo, en el caso de la ERCUE, su objetivo principal radica en la medición de los patrones de consumo energético del segmento residencial a nivel nacional. Esto permite evaluar los patrones de consumo de diversas fuentes de energía y su relación con las condiciones socioeconómicas de los hogares. En este sentido, para aprovechar esta información de manera adecuada, es necesario primero realizar una caracterización de los residuos, lo cual implica identificar sus componentes y características específicas.

Así pues, presentamos dos análisis estadísticos robustos para validar nuestras hipótesis de investigación y metodología de trabajo. El primero está relacionado a la implementación de una regresión lineal múltiple con la finalidad de validar la significancia del consumo de

combustibles en hogares con relación a la generación de residuos. Por otro lado, implementamos un modelo predictivo basado en los algoritmos de *random forest* debido a su aplicabilidad en clases tabulares y robustez ante diferentes tipos de variables. Asimismo, la optimización en la selección de variables, la mejora de la calidad y disponibilidad de datos, y la consideración de la variabilidad a nivel individual son elementos relevantes para avanzar hacia una gestión más efectiva y sostenible de los RSM. A continuación, responderemos a las preguntas de investigación propuestas al principio de esta tesis, utilizando nuestro desarrollo conceptual y los resultados de los análisis realizados.

6.1 PI – 01: ¿Cómo podemos caracterizar a los patrones de consumo, medios y condiciones de vida de los habitantes del hogar en función a los residuos sólidos generados?

Respondimos esta pregunta realizando una revisión crítica semisistemática de la literatura disponible durante las últimas dos décadas. De esta manera, nos enfocamos en poder entender el vínculo existente entre las distintas consideraciones antropogénicas que caracterizan al habitante del hogar frente a la generación de RSM. Exploramos distintas sinergias asociadas al comportamiento en el hogar y la influencia de estos en el cómputo global de RSM. También revisamos las distintas formas de caracterizar a una población en función de las fuentes de información disponible (p. ej., censos, bases de datos de acceso abierto, etc.) y evidenciamos las limitantes al respecto. Además, validamos que la identificación de variables con fines predictivos está influenciada, en su gran mayoría, por el alcance temporal de los estudios y disponibilidad de información. Concluimos que podemos describir a la población en tres dimensiones: socioculturales, económicas y ambientales. Esto nos permitió confirmar la Hipótesis 1², ya que las dimensiones socioculturales, económicas y ambientales ofrecen

² H – 1: Las dimensiones sociales, económicas y ambientales ofrecen robustez y significancia para caracterizar los patrones de consumo, medios y condiciones de vida de los residentes.

robustez y significancia para caracterizar los patrones de consumo, medios y condiciones de vida de los residentes.

Por otro lado, esta revisión crítica permitió también visibilizar limitaciones importantes asociadas a una gestión integral de RSM relacionadas a la escasez de fuentes alternativas de información y a la falta de validación de variables. Esto enfatiza la necesidad de un enfoque más riguroso y estandarizado en la selección de variables y la recopilación de datos para futuras investigaciones. Además, la mejora en la calidad y disponibilidad de la data se torna esencial, ya que la falta de información en ciertos municipios dificulta la gestión de residuos en esas áreas. La necesidad de fuentes confiables y robustas también ha sido reconocida en otros estudios con fines predictivos o exploratorios. Aunque lo que proponemos en esta investigación es una salida viable para la predicción, aún es necesario dedicar más esfuerzos para mejorar los métodos y estrategias de identificación de variables potencialmente capaces de predecir los RSM.

Dentro de las variables analizadas, se ha identificado que la edad, el nivel educativo, la cantidad de habitantes por hogar y el gasto promedio mensual por persona son factores significativos en la generación de RSM. La edad, como evidencia la literatura revisada, está directamente relacionada con los patrones de consumo, siendo las poblaciones más jóvenes las que tienden a generar más residuos. Por otro lado, en cuanto al nivel educativo, se observa que las comunidades con niveles educativos más altos tienden a ser más conscientes y a practicar una gestión más responsable de los residuos. Además, la cantidad de habitantes por hogar tiene un impacto notable en la generación de RSM, siendo los hogares más poblados los que generan una mayor cantidad de residuos. Asimismo, la variable del gasto promedio mensual por persona refleja cómo las condiciones económicas también inciden en la generación de RSM, ya que un mayor poder adquisitivo puede estar asociado a un consumo más elevado y, por ende, a una mayor producción de residuos.

6.2 PI – 02: ¿Cómo podemos identificar qué tipo de variables presentan correlación con el proceso de generación de RSM?

Sugerimos que las aproximaciones metodológicas para identificar variables predictoras empiecen con un análisis exploratorio causal para que, de esta forma, se pueda captar las relaciones propias del caso de estudio. En este proyecto, hemos desarrollado un análisis exploratorio causal en torno a las dimensiones anteriormente descritas para poder identificar la pertenencia de inclusión de determinadas variables dentro del set predictivo. Posteriormente, con las variables previamente identificadas, se llevó a cabo un análisis de regresión lineal múltiple con el fin de evaluar su significancia respectiva.

La utilización de la regresión lineal múltiple, en este contexto, se sustenta en su capacidad para modelar relaciones lineales entre variables, permitiendo así explorar y cuantificar la contribución relativa de cada variable independiente al resultado deseado. Aunque se admite la simplificación inherente de este enfoque respecto a la complejidad real de las interacciones, la regresión lineal múltiple permite identificar tendencias y patrones generales que pueden ser relevantes en la generación de RSM. De este modo, pudimos corroborar nuestra Hipótesis 2³, ya que fue posible identificar si existe o no significancia al relacionar la variable elegida y la cantidad de residuo generado a través de un análisis de regresión lineal múltiple.

Esta investigación proporciona un primer intento de establecer una metodología conceptual para la identificación de variables predictoras, presentando al análisis exploratorio causal y a la regresión lineal múltiple como herramientas adecuadas para facilitar la inclusión o exclusión de una determinada variable. Dada la complejidad de los patrones de consumo, es importante evaluar constantemente el dinamismo económico externo que muchas veces rige el comportamiento en el interior de una comunidad.

³ H – 2: Un análisis de regresión lineal múltiple permitirá identificar si existe o no significancia al relacionar la variable elegida y la cantidad de residuo generado.

6.3 PI – 03: ¿Cómo podemos medir indirectamente la generación de residuos sólidos municipales en un contexto de zonas urbanas?

En función de los resultados obtenidos de las PI – 01⁴ y PI – 02⁵, hemos notado una anomia institucional en lo que respecta a recolección de información. Esto implica que las mediciones asociadas a la generación de residuos sólidos domiciliarios o municipales a un nivel de observación equivalente a vivienda es nulo o escaso. En ese sentido, en vista de que la etiqueta principal (p. ej., GPC de RSM) no es accesible con regularidad o facilidad, se hace necesario la inclusión de variables indirectas que faciliten su comprensión. La inclusión de este tipo de variables sumadas a la utilización de datos asociados a los patrones de consumo y características antropogénicas de los residentes ofrecen un panorama completo para estimar la generación actual de RSM. En ese sentido, analizando la influencia de factores culturales y comportamentales que puedan afectar la cantidad de residuos generados por habitante, el consumo de combustible erige como factor relevante a analizar debido a la sinergia evidenciada en el análisis exploratorio causal y sustentada en la significancia obtenida en el análisis de regresión. Esta aproximación proporciona una alternativa eficaz cuando las mediciones directas son difíciles de implementar o costosas, permitiendo una evaluación pragmática y viable de la generación de residuos sólidos en contextos urbanos. Por consiguiente, pudimos demostrar la Hipótesis 3⁶, ya que expusimos que el uso de combustible para cocinar en hogares se torna como un indicador sólido y confiable para predecir de manera indirecta la generación de RSM.

⁴ PI – 01: ¿Cómo podemos caracterizar a los patrones de consumo, medios y condiciones de vida de los habitantes del hogar en función a los residuos sólidos generados?

⁵ PI – 02: ¿Cómo podemos identificar qué tipo de variables presentan correlación con el proceso de generación de RSM?

⁶ H – 3: El uso de combustible para cocinar se presenta como un indicador robusto y fiable para medir indirectamente el residuo generado.

6.4 PI – 04: ¿Cómo podemos integrar las variables recopiladas en un modelo de aprendizaje automático destinado a la predicción de residuos sólidos?

Tomando como insumo los alcances provistos por la PI – 03⁷, se realizó una revisión holística de las variables identificadas, evaluando su relevancia y contribución potencial a la predicción de la generación de RSM. Posteriormente, llevamos a cabo un proceso exploratorio de los datos para abordar posibles problemas de confiabilidad de la información y fuentes disponibles. En ese sentido, considerando la ausencia de información relevante en una única base de datos, se tuvo que recurrir a otras dos fuentes de datos distintas; y así con ello se trabajó en la construcción de una fuente de información unificada que sirva como insumo para el modelo predictivo.

La integración de múltiples fuentes de datos —aunque valiosa— conlleva incertidumbre y simplificaciones que deben ser consideradas cuidadosamente al interpretar los resultados. En este contexto, se decidió emplear el algoritmo *random forest* como modelo de aprendizaje automático. La elección de *random forest* se basó en su capacidad para manejar conjuntos de datos complejos y heterogéneos, así como su habilidad para manejar múltiples variables predictoras. Este modelo demostró ser efectivo en problemas de predicción, proporcionando resultados robustos y generalizables. Por lo tanto, validamos nuestra Hipótesis 4⁸ dado que el *random forest* fue el modelo que mejor ajuste tuvo a los datos recolectados. Además, se proporcionaron perspectivas para futuras mejoras y refinamientos del modelo, asegurando así una integración sólida y efectiva de las variables en el proceso de predicción.

⁷ PI – 03: ¿Cómo podemos medir indirectamente la generación de RSM en un contexto de zonas urbanas?

⁸ H – 04: Dada la naturaleza de las variables, un modelo de random forest, se ajusta mejor a los objetivos de este estudio.

6.5 Recomendaciones

Para establecer una gestión eficiente de los RSM, es necesario contar con información disponible y actualizada. En ese sentido, se recomienda que el Estado, en representación de todas las agencias gubernamentales, repotencie las estrategias de recolección de información a través de programas destinados a desarrollar al menos un estudio de caracterización de RSM de manera anualizada. De este modo, la fracción orgánica domiciliaria reflejará a mayor detalle los patrones de consumo, permitiendo a futuras investigaciones ampliar el alcance de estudio y considerar otras fuentes generadoras de este tipo de residuos.

En cuanto a la recolección de información, se sugiere que las encuestas capturen un mayor número de observaciones a lo largo de cada distrito para poder conseguir una mayor representatividad de las muestras y no perder información. Además, se hace hincapié en la necesidad de una validación detallada de variables y la exploración de alternativas para la predicción. Al pronosticar la generación de residuos, es importante considerar las actividades diarias de la población y realizar una cuidadosa selección de variables predictoras que describan el contexto y puedan prever condiciones futuras.

Por otro lado, se subraya la importancia de llevar a cabo un análisis temporal y territorial, y se alienta a los investigadores a determinar un conjunto específico de variables socioculturales, ambientales y económicas, siguiendo la metodología propuesta en esta tesis. Por último, se recomienda a las autoridades locales que desarrollen políticas y regulaciones específicas basadas en estrategias locales, lo que garantizará una gestión de residuos más efectiva y adaptada a las necesidades propias de la comunidad.

REFERENCIAS

- Abbasi, M., & Hanandeh, A. El. (2016). Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Management*, 56, 13–22. <https://doi.org/10.1016/j.wasman.2016.05.018>
- Abbasi, M., Rastgoo, M. N., & Nakisa, B. (2018). Monthly and seasonal modeling of municipal waste generation using radial basis function neural network. *Environmental Progress & Sustainable Energy*, 38(3), e13033. <https://doi.org/10.1002/ep.13033>
- Abdallah, M., Arab, M., Shabib, A., El-Sherbiny, R., & El-Sheltawy, S. (2020). Characterization and sustainable management strategies of municipal solid waste in Egypt. *Clean Technologies and Environmental Policy*, 22(6), 1371–1383. <https://doi.org/10.1007/s10098-020-01877-0>
- Abdallah, M., Talib, M. A., Feroz, S., Nasir, Q., Abdalla, H., & Mahfood, B. (2020). Artificial intelligence applications in solid waste management: A systematic research review. *Waste Management*, 109, 231–246. <https://doi.org/10.1016/j.wasman.2020.04.057>
- Abdel-Shafy, H. I., & Mansour, M. S. M. (2018). Solid waste issue: Sources, composition, disposal, recycling, and valorization. *Egyptian Journal of Petroleum*, 27(4), 1275–1290. <https://doi.org/10.1016/j.ejpe.2018.07.003>
- Abdella Ahmed, A. K., Ibraheem, A. M., & Abd-Ellah, M. K. (2022). Forecasting of municipal solid waste multi-classification by using time-series deep learning depending on the living standard. *Results in Engineering*, 16(July), 100655. <https://doi.org/10.1016/j.rineng.2022.100655>
- Abdoli, M A, Nezhad, M. F., Sede, R. S., & Behboudian, S. (2011). Longterm forecasting of solid waste generation by the artificial neural networks. *Environmental Progress &*

Sustainable Energy, 33(3), 676–680. <https://doi.org/10.1002/ep>

Abdoli, M Ali, Nezhad, M. F., & Behboudian, S. (2011). Multivariate econometric approach for solid waste generation modeling: Impact of climate factors. *Environmental Engineering Science*, 28(9), 627–633. <https://doi.org/10.1089/ees.2010.0234>

Abuín, J. M. R. (2008). Multiple linear regression. *Instituto de Economía y Geografía*, 150(10), 22–23.

Agovino, M., Cerciello, M., Musella, G., & Garofalo, A. (2024). European waste management regulations and the transition towards circular economy. A shift-and-share analysis. *Journal of Environmental Management*, 354(February), 120423. <https://doi.org/10.1016/j.jenvman.2024.120423>

Aja, O. C., & Al-Kayiem, H. H. (2014). Review of municipal solid waste management options in Malaysia, with an emphasis on sustainable waste-to-energy options. *Journal of Material Cycles and Waste Management*, 16(4), 693–710. <https://doi.org/10.1007/s10163-013-0220-z>

Al-Khatib, I. A., Monou, M., Zahra, A. S. F. A., Shaheen, H. Q., & Kassinos, D. (2010). Solid waste characterization, quantification and management practices in developing countries. A case study: Nablus district - Palestine. *Journal of Environmental Management*, 91(5), 1131–1138. <https://doi.org/10.1016/j.jenvman.2010.01.003>

Albert, A. O., & Olutayo, F. S. (2021). Cultural dimensions of environmental problems: A critical overview of solid waste generation and management in Nigeria. *American International Journal of Multidisciplinary Scientific Research*, 8(1), 1–15. <https://doi.org/10.46281/aijmsr.v8i1.1110>

Ananda, J., Karunasena, G. G., Kansal, M., Mitsis, A., & Pearson, D. (2023). Quantifying the

- effects of food management routines on household food waste. *Journal of Cleaner Production*, 391(January 2022), 136230. <https://doi.org/10.1016/j.jclepro.2023.136230>
- Andrade, C. (2019). The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives. *Indian Journal of Psychological Medicine*, 41(3), 210–215. https://doi.org/10.4103/IJPSYM.IJPSYM_193_19
- Antanasijević, D., Pocajt, V., Popović, I., Redžić, N., & Ristić, M. (2013). The forecasting of municipal waste generation using artificial neural networks and sustainability indicators. *Sustainability Science*, 8(1), 37–46. <https://doi.org/10.1007/s11625-012-0161-9>
- APEIM. (2019). *Socioeconomic levels*. <https://www.ptonline.com/articles/how-to-get-better-mfi-results>
- Assef, F. M., Steiner, M. T. A., & Lima, E. P. de. (2022). A review of clustering techniques for waste management. *Heliyon*, 8(1), e08784. <https://doi.org/10.1016/j.heliyon.2022.e08784>
- Azadi, S., & Karimi-Jashni, A. (2016). Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of Fars province, Iran. *Waste Management*, 48, 14–23. <https://doi.org/10.1016/j.wasman.2015.09.034>
- Azevedo, B. D., Scavarda, L. F., Caiado, R. G. G., & Fuss, M. (2021). Improving urban household solid waste management in developing countries based on the German experience. *Waste Management*, 120, 772–783. <https://doi.org/10.1016/j.wasman.2020.11.001>
- Bedoya-Perales, N. S., & Magro, G. P. D. (2021). Quantification of food losses and waste in peru: A mass flow analysis along the food supply chain. *Sustainability*, 13(5), 1–15.

<https://doi.org/10.3390/su13052807>

Beigl, P., Lebersorger, S., & Salhofer, S. (2008). Modelling municipal solid waste generation: A review. *Waste Management*, 28(1), 200–214.

<https://doi.org/10.1016/j.wasman.2006.12.011>

Blasenbauer, D., Lipp, A.-M., Fellner, J., Tischberger-Aldrian, A., Stipanović, H., & Lederer, J. (2024). Recovery of plastic packaging from mixed municipal solid waste. A case study from Austria. *Waste Management*, 180(February), 9–22.

<https://doi.org/10.1016/j.wasman.2024.02.040>

Buenrostro, O., Bocco, G., & Vence, J. (2001). Forecasting generation of urban solid waste in developing countries - A case study in Mexico. *Journal of the Air and Waste Management Association*, 51(1), 86–93. <https://doi.org/10.1080/10473289.2001.10464258>

Burkart, N., & Huber, M. F. (2021). A Survey on the Explainability of Supervised Machine Learning. *Journal of Artificial Intelligence Research*, 70, 245–317.

<https://doi.org/10.1613/jair.1.12228>

CalRecycle. (2024). *Waste Characterization*.

<https://www2.calrecycle.ca.gov/WasteCharacterization/>

Cárdenas-Mamani, Ú., Kahhat, R., & Vázquez-Rowe, I. (2022). District-level analysis for household-related energy consumption and greenhouse gas emissions: A case study in Lima, Peru. *Sustainable Cities and Society*, 77. <https://doi.org/10.1016/j.scs.2021.103572>

Caruso, G., & Gattone, S. A. (2019). Waste management analysis in developing countries through unsupervised classification of mixed data. *Social Sciences*, 8(6).

<https://doi.org/10.3390/socsci8060186>

Cervantes, D. E. T., Martínez, A. L., Hernández, M. C., & de Cortázar, A. L. G. (2018). Using

- indicators as a tool to evaluate municipal solid waste management: A critical review. *Waste Management*, 80, 51–63. <https://doi.org/10.1016/j.wasman.2018.08.046>
- Chand Malav, L., Yadav, K. K., Gupta, N., Kumar, S., Sharma, G. K., Krishnan, S., Rezania, S., Kamyab, H., Pham, Q. B., Yadav, S., Bhattacharyya, S., Yadav, V. K., & Bach, Q.-V. (2020). A review on municipal solid waste as a renewable source for waste-to-energy project in India: Current practices, challenges, and future opportunities. *Journal of Cleaner Production*, 277, 123227. <https://doi.org/10.1016/j.jclepro.2020.123227>
- Chang, N. Bin, & Davila, E. (2008). Municipal solid waste characterizations and management strategies for the Lower Rio Grande Valley, Texas. *Waste Management*, 28(5), 776–794. <https://doi.org/10.1016/j.wasman.2007.04.002>
- Chhay, L., Reyad, M. A. H., Suy, R., Islam, M. R., & Mian, M. M. (2018). Municipal solid waste generation in China: influencing factor analysis and multi-model forecasting. *Journal of Material Cycles and Waste Management*, 20(3), 1761–1770. <https://doi.org/10.1007/s10163-018-0743-4>
- Çinar, Z. M., Nuhu, A. A., Zeeshan, Q., Korhan, O., Asmael, M., & Safaei, B. (2020). Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. *Sustainability (Switzerland)*, 12(19). <https://doi.org/10.3390/su12198211>
- Cristóbal, J., Vázquez-Rowe, I., Margallo, M., Ita-Nagy, D., Ziegler-Rodriguez, K., Laso, J., Ruiz-Salmón, I., Kahhat, R., & Aldaco, R. (2022). Climate change mitigation potential of transitioning from open dumpsters in Peru: Evaluation of mitigation strategies in critical dumpsites. *Science of The Total Environment*, 846, 157295. <https://doi.org/10.1016/j.scitotenv.2022.157295>
- Cubillos, M. (2020). Multi-site household waste generation forecasting using a deep learning approach. *Waste Management*, 115, 8–14. <https://doi.org/10.1016/j.wasman.2020.06.046>

- Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). *Random Forests BT - Ensemble Machine Learning: Methods and Applications* (C. Zhang & Y. Ma (eds.); pp. 157–175). Springer New York. https://doi.org/10.1007/978-1-4419-9326-7_5
- Daly, H. E., & Farley, J. (2011). *Ecological Economics: Principles and Applications* (Issue July). Island Press.
- Deus, R. M., Mele, F. D., Bezerra, B. S., & Battistelle, R. A. G. (2020). A municipal solid waste indicator for environmental impact: Assessment and identification of best management practices. *Journal of Cleaner Production*, 242, 118433. <https://doi.org/10.1016/j.jclepro.2019.118433>
- Dou, Z., & Toth, J. D. (2021). Global primary data on consumer food waste: Rate and characteristics – A review. *Resources, Conservation and Recycling*, 168(June 2020), 105332. <https://doi.org/10.1016/j.resconrec.2020.105332>
- Durand, M. (2011). La Gestión de los Residuos Sólidos en los Países en Desarrollo: ¿Cómo obtener beneficios de las dificultades actuales? *Espacio y Desarrollo*, 23, 115–130. <http://revistas.pucp.edu.pe/index.php/espacioydesarrollo/article/view/3502>
- Dyson, B., & Chang, N. Bin. (2005). Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling. *Waste Management*, 25(7), 669–679. <https://doi.org/10.1016/j.wasman.2004.10.005>
- EEA. (2022). *Early warning assessment related to the 2025 targets for municipal waste and packaging waste*.
- Elsevier, S. (2024). *Scopus Web site*. <https://www.scopus.com>
- EPA. (2020). *The Best Practices for Solid Waste Management: A Guide for Decision-Makers in Developing Countries*. <https://www.epa.gov/sites/production/files/2021->

02/documents/swm_guide-spanish-reducedfilesize_pubnumber_october.pdf

Espinoza Pérez, L., Ziegler-Rodríguez, K., Espinoza Pérez, A. T., Vásquez, Ó. C., & Vázquez-Rowe, I. (2021). Closing the gap in the municipal solid waste management between metropolitan and regional cities from developing countries: A life cycle assessment approach. *Waste Management*, *124*, 314–324. <https://doi.org/10.1016/j.wasman.2021.02.020>

FAO. (2011). *Global food losses and food waste – Extent, causes and prevention*.

FAO. (2014). Global initiative on food loss and waste reduction. Definitional Framework of Food Loss. In *Food And Agriculture Organization Of The United Nations; FAO*. http://www.fao.org/fileadmin/user_upload/save-food/PDF/FLW_Definition_and_Scope_2014.pdf

Frank, K. A., Lin, Q., Xu, R., Maroulis, S., & Mueller, A. (2022). Quantifying the robustness of causal inferences: Sensitivity analysis for pragmatic social science. *Social Science Research*, *110*(May 2022), 102815. <https://doi.org/10.1016/j.ssresearch.2022.102815>

Garcia-Herrero, I., Hoehn, D., Margallo, M., Laso, J., Bala, A., Batlle-Bayer, L., Fullana, P., Vazquez-Rowe, I., Gonzalez, M. J., Durá, M. J., Sarabia, C., Abajas, R., Amo-Setien, F. J., Quiñones, A., Irabien, A., & Aldaco, R. (2018). On the estimation of potential food waste reduction to support sustainable production and consumption policies. *Food Policy*, *80*(November 2017), 24–38. <https://doi.org/10.1016/j.foodpol.2018.08.007>

Gaviria-Cuevas, J. F., Soto-Paz, J., Manyoma-Velasquez, P. C., & Torres-Lozada, P. (2019). Trends in research on the supply chain management of municipal solid waste. *Información Tecnológica*, *30*(4), 147–154. <https://doi.org/10.4067/s0718-07642019000400147>

Giampietro, M., & Mayumi, K. (2018). Unraveling the complexity of the Jevons Paradox: The

- link between innovation, efficiency, and sustainability. *Frontiers in Energy Research*, 6(APR), 1–13. <https://doi.org/10.3389/fenrg.2018.00026>
- González-Santana, R. A., Blesa, J., Frígola, A., & Esteve, M. J. (2022). Dimensions of household food waste focused on family and consumers. *Critical Reviews in Food Science and Nutrition*, 62(9), 2342–2354. <https://doi.org/10.1080/10408398.2020.1853033>
- Graham-Rowe, E., Jessop, D. C., & Sparks, P. (2014). Identifying motivations and barriers to minimising household food waste. *Resources, Conservation and Recycling*, 84, 15–23. <https://doi.org/10.1016/j.resconrec.2013.12.005>
- Grazhdani, D. (2016). Assessing the variables affecting on the rate of solid waste generation and recycling: An empirical analysis in Prespa Park. *Waste Management*, 48, 3–13. <https://doi.org/10.1016/j.wasman.2015.09.028>
- Gu, B., Jiang, S., Wang, H., Wang, Z., Jia, R., Yang, J., He, S., & Cheng, R. (2016). Characterization, quantification and management of China's municipal solid waste in spatiotemporal distributions: A review. *Waste Management*, 61, 67–77. <https://doi.org/10.1016/j.wasman.2016.11.039>
- Guo, H. nan, Wu, S. biao, Tian, Y. jie, Zhang, J., & Liu, H. tao. (2021). Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. *Bioresource Technology*, 319(July 2020), 124114. <https://doi.org/10.1016/j.biortech.2020.124114>
- Han, Z., Liu, Y., Zhong, M., Shi, G., Li, Q., Zeng, D., Zhang, Y., Fei, Y., & Xie, Y. (2017). Influencing factors of domestic waste characteristics in rural areas of developing countries. *Waste Management*, 72, 45–54. <https://doi.org/10.1016/j.wasman.2017.11.039>
- Hannan, M. A., Lipu, M. S. H., Akhtar, M., Begum, R. A., Mamun, M. A. Al, Hussain, A.,

- Mia, M. S., & Basri, H. (2020). Solid waste collection optimization objectives, constraints, modeling approaches, and their challenges toward achieving sustainable development goals. *Journal of Cleaner Production*, 277, 123557. <https://doi.org/10.1016/j.jclepro.2020.123557>
- Hemali-Arosha, N., De Alwis, A. A. P., Wijesundara, M., & Edirisinghe, L. G. L. M. (2024). A method for determining the recycling value of unprocessed municipal solid waste in one cubic meter waste composition analysis technique. *MethodsX*, 12(March). <https://doi.org/10.1016/j.mex.2024.102626>
- Hoehn, D., Margallo, M., Laso, J., Fernández-Ríos, A., Ruiz-Salmón, I., & Aldaco, R. (2022). Energy Systems in the Food Supply Chain and in the Food Loss and Waste Valorization Processes: A Systematic Review. *Energies*, 15(6). <https://doi.org/10.3390/en15062234>
- Hoehn, D., Vázquez-Rowe, I., Kahhat, R., Margallo, M., Laso, J., Fernández-Ríos, A., Ruiz-Salmón, I., & Aldaco, R. (2023). A critical review on food loss and waste quantification approaches: Is there a need to develop alternatives beyond the currently widespread pathways? *Resources, Conservation and Recycling*, 188, 106671. <https://doi.org/10.1016/j.resconrec.2022.106671>
- Hoornweg, D., & Bhada-Tata, P. (2012). What a waste: a global review of solid waste management. In *Urban development series knowledge papers* (Vol. 15). <https://openknowledge.worldbank.org/handle/10986/17388%0Ahttps://openknowledge.worldbank.org/handle/10986/17388%0Ahttps://openknowledge.worldbank.org/handle/10986/17388%0Awww.worldbank.org/urban>
- Ibarra-Esparza, F. E., González-López, M. E., Ibarra-Esparza, J., Lara-Topete, G. O., Senés-Guerrero, C., Cansdale, A., Forrester, S., Chong, J. P. J., & Gradilla-Hernández, M. S. (2023). Implementation of anaerobic digestion for valorizing the organic fraction of

- municipal solid waste in developing countries: Technical insights from a systematic review. *Journal of Environmental Management*, 347(August).
<https://doi.org/10.1016/j.jenvman.2023.118993>
- INEI. (2017). *Inei - Redatam Censos 2017*. <http://censos2017.inei.gob.pe/redatam/>
- INEI. (2021). *Performance indicators of budget programs 2020 - Annual results* (p. 441).
<http://proyecto.inei.gob.pe/enapres/wp-content/uploads/2021/04/ENAPRES-Indicadores-de-Programas-Presupuestales-2020.pdf>
- Iordache, M., De Keukelaere, L., Moelans, R., Landuyt, L., Moshtaghi, M., Corradi, P., & Knaeps, E. (2022). Targeting Plastics: Machine Learning Applied to Litter Detection in Aerial Multispectral Images. *Remote Sensing*, 14(22), 5820.
<https://doi.org/10.3390/rs14225820>
- Isah, S., & Ozbay, G. (2020). Valorization of Food Loss and Wastes: Feedstocks for Biofuels and Valuable Chemicals. *Frontiers in Sustainable Food Systems*, 4(June).
<https://doi.org/10.3389/fsufs.2020.00082>
- Izquierdo-Horna, L., Damazo, M., & Yanayaco, D. (2022). Identification of urban sectors prone to solid waste accumulation: A machine learning approach based on social indicators. *Computers, Environment and Urban Systems*, 96, 101834.
<https://doi.org/10.1016/j.compenvurbsys.2022.101834>
- Izquierdo-Horna, L., Kahhat, R., & Vázquez-Rowe, I. (2022). Reviewing the influence of sociocultural, environmental and economic variables to forecast municipal solid waste (MSW) generation. *Sustainable Production and Consumption*, 33, 809–819.
<https://doi.org/10.1016/j.spc.2022.08.008>
- Izquierdo-Horna, L., Kahhat, R., & Vázquez-Rowe, I. (2023). Identification of variables

related to household solid waste generation forecasting in residential areas. *19th International Symposium on Waste Management, Resource Recovery and Sustainable Landfilling*.

Izquierdo-Horna, L., Kahhat, R., & Vázquez-Rowe, I. (2024). Unveiling the energy consumption - food waste nexus in households : A focus on key predictors of food waste generation. *Journal of Material Cycles and Waste Management*.
<https://doi.org/10.1007/s10163-024-01946-2>

Johnson, N. E., Ianiuk, O., Cazap, D., Liu, L., Starobin, D., Dobler, G., & Ghandehari, M. (2017). Patterns of waste generation: A gradient boosting model for short-term waste prediction in New York City. *Waste Management*, 62, 3–11.
<https://doi.org/10.1016/j.wasman.2017.01.037>

Jordán, R., Riffo, L., & Prado, A. (2017). Sustainable development, urbanization and inequality in Latin America and the Caribbean: dynamics and challenges for structural change. *Cepal*, October, 427. <https://www.cepal.org/es/publicaciones/42141-desarrollo-sostenible-urbanizacion-desigualdad-america-latina-caribe-dinamicas>

Kan, M., & Miller, S. A. (2022). Environmental impacts of plastic packaging of food products. *Resources, Conservation and Recycling*, 180(November 2021), 106156.
<https://doi.org/10.1016/j.resconrec.2022.106156>

Kannangara, M., Dua, R., Ahmadi, L., & Bensebaa, F. (2018). Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste Management*, 74, 3–15. <https://doi.org/10.1016/j.wasman.2017.11.057>

Karadimas, N. V., & Orsoni, A. (2006). Municipal solid waste generation modelling based on fuzzy logic. *20th European Conference on Modelling and Simulation: Modelling Methodologies and Simulation Key Technologies in Academia and Industry, ECMS 2006*,

5(Cd), 309–314. <https://doi.org/10.7148/2006-0309>

Karak, T., Bhagat, R. M., & Bhattacharyya, P. (2012). Municipal solid waste generation, composition, and management: The world scenario. *Critical Reviews in Environmental Science and Technology*, 42(15), 1509–1630. <https://doi.org/10.1080/10643389.2011.569871>

Kaur, P., Kaur, G. J., Routray, W., Rahimi, J., Nair, G. R., & Singh, A. (2021). Recent advances in utilization of municipal solid waste for production of bioproducts: A bibliometric analysis. *Case Studies in Chemical and Environmental Engineering*, 4(October), 100164. <https://doi.org/10.1016/j.cscee.2021.100164>

Kaza, S., Yao, L. C., Bhada-Tata, P., & Van Woerden, F. (2018). *What a Waste 2.0 : A Global Snapshot of Solid Waste Management to 2050*. <https://openknowledge.worldbank.org/handle/10986/30317>

Keser, S., Duzgun, S., & Aksoy, A. (2012). Application of spatial and non-spatial data analysis in determination of the factors that impact municipal solid waste generation rates in Turkey. *Waste Management*, 32(3), 359–371. <https://doi.org/10.1016/j.wasman.2011.10.017>

Khair, H., Rachman, I., & Matsumoto, T. (2019). Analyzing household waste generation and its composition to expand the solid waste bank program in Indonesia: a case study of Medan City. *Journal of Material Cycles and Waste Management*, 21(4), 1027–1037. <https://doi.org/10.1007/s10163-019-00840-6>

Khan, M., & Burney, F. (1989). Forecasting solid waste composition - An important consideration in resource recovery and recycling. *Resources, Conservation and Recycling*, 3(1), 1–17. [https://doi.org/10.1016/0921-3449\(89\)90010-4](https://doi.org/10.1016/0921-3449(89)90010-4)

- Khan, S., Anjum, R., Raza, S. T., Ahmed Bazai, N., & Ihtisham, M. (2022). Technologies for municipal solid waste management: Current status, challenges, and future perspectives. *Chemosphere*, 288(P1), 132403. <https://doi.org/10.1016/j.chemosphere.2021.132403>
- Kladnik, V., Dworak, S., & Schwarzböck, T. (2024). Composition of public waste - a case study from Austria. *Waste Management*, 178(June 2023), 210–220. <https://doi.org/10.1016/j.wasman.2024.02.031>
- Kolekar, K. A., Hazra, T., & Chakrabarty, S. N. (2016). A Review on Prediction of Municipal Solid Waste Generation Models. *Procedia Environmental Sciences*, 35, 238–244. <https://doi.org/10.1016/j.proenv.2016.07.087>
- Kollikkathara, N., Feng, H., & Yu, D. (2010). A system dynamic modeling approach for evaluating municipal solid waste generation, landfill capacity and related cost management issues. *Waste Management*, 30(11), 2194–2203. <https://doi.org/10.1016/j.wasman.2010.05.012>
- Komakech, A. J., Banadda, N. E., Kinobe, J. R., Kasisira, L., Sundberg, C., Gebresenbet, G., & Vinnerås, B. (2014). Characterization of municipal waste in Kampala, Uganda. *Journal of the Air and Waste Management Association*, 64(3), 340–348. <https://doi.org/10.1080/10962247.2013.861373>
- Kumar, A., & Samadder, S. R. (2017). An empirical model for prediction of household solid waste generation rate – A case study of Dhanbad, India. *Waste Management*, 68, 3–15. <https://doi.org/10.1016/j.wasman.2017.07.034>
- Kundariya, N., Mohanty, S. S., Varjani, S., Ngo, H. H., Wong, J. W. C., Taherzadeh, M. J., Chang, J. S., Ng, H. Y., Kim, S. H., & Bui, X. T. (2021). A review on integrated approaches for municipal solid waste for environmental and economical relevance: Monitoring tools, technologies, and strategic innovations. *Bioresource Technology*,

342(May), 125982. <https://doi.org/10.1016/j.biortech.2021.125982>

Lagerkvist, A., & Dahlén, L. (2012). Solid Waste Generation and Characterization. In R. A. Meyers (Ed.), *Encyclopedia of Sustainability Science and Technology* (pp. 10000–10013). Springer New York. https://doi.org/10.1007/978-1-4419-0851-3_110

Lakhout, A., Shaban, M., Alatawi, A., Abbas, S. Y. H., Asiri, E., Al Juhni, T., & Elsayy, M. (2023). Machine-learning approaches in geo-environmental engineering: Exploring smart solid waste management. *Journal of Environmental Management*, 330(November 2022), 117174. <https://doi.org/10.1016/j.jenvman.2022.117174>

Laso, J., Margallo, M., García-Herrero, I., Fullana, P., Bala, A., Gazulla, C., Poletini, A., Kahhat, R., Vázquez-Rowe, I., Irabien, A., & Aldaco, R. (2018). Combined application of Life Cycle Assessment and linear programming to evaluate food waste-to-food strategies: Seeking for answers in the nexus approach. *Waste Management*, 80, 186–197. <https://doi.org/10.1016/j.wasman.2018.09.009>

Lebersorger, S., & Beigl, P. (2011). Municipal solid waste generation in municipalities: Quantifying impacts of household structure, commercial waste and domestic fuel. *Waste Management*, 31(9–10), 1907–1915. <https://doi.org/10.1016/j.wasman.2011.05.016>

Lella, J., Mandla, V. R., & Zhu, X. (2017). Solid waste collection/transport optimization and vegetation land cover estimation using Geographic Information System (GIS): A case study of a proposed smart-city. *Sustainable Cities and Society*, 35(August), 336–349. <https://doi.org/10.1016/j.scs.2017.08.023>

Li, S., & Skelly, S. (2023). Physicochemical properties and applications of biochars derived from municipal solid waste: A review. *Environmental Advances*, 13(May), 100395. <https://doi.org/10.1016/j.envadv.2023.100395>

- Li, Z. S., Fu, H. Z., & Qu, X. Y. (2011). Estimating municipal solid waste generation by different activities and various resident groups: A case study of Beijing. *Science of the Total Environment*, 409(20), 4406–4414. <https://doi.org/10.1016/j.scitotenv.2011.07.018>
- Liu, B., Zhang, L., & Wang, Q. (2021). Demand gap analysis of municipal solid waste landfill in Beijing: Based on the municipal solid waste generation. *Waste Management*, 134(April), 42–51. <https://doi.org/10.1016/j.wasman.2021.08.007>
- Liu, J., Li, Q., Gu, W., & Wang, C. (2019). The impact of consumption patterns on the generation of municipal solid waste in China: Evidences from provincial data. *International Journal of Environmental Research and Public Health*, 16(10), 1–19. <https://doi.org/10.3390/ijerph16101717>
- Ma, J., & Hipel, K. W. (2016). Exploring social dimensions of municipal solid waste management around the globe – A systematic literature review. *Waste Management*, 56, 3–12. <https://doi.org/10.1016/j.wasman.2016.06.041>
- Magrinho, A., Didelet, F., & Semiao, V. (2006). Municipal solid waste disposal in Portugal. *Waste Management*, 26(12), 1477–1489. <https://doi.org/10.1016/j.wasman.2006.03.009>
- Margallo, M., Ziegler-rodriguez, K., Vázquez-Rowe, I., Aldaco, R., Irabien, Á., & Kahhat, R. (2019). Enhancing waste management strategies in Latin America under a holistic environmental assessment perspective: A review for policy support. *Science of the Total Environment*, 689, 1255–1275. <https://doi.org/10.1016/j.scitotenv.2019.06.393>
- MINAM. (2019a). *Guide for the characterization of municipal solid waste*. https://cdn.www.gob.pe/uploads/document/file/523785/Guía_para_la_caracterización_rm-29012020__1_.pdf?v=1581976231
- MINAM. (2019b). *SIGERSOL*. <https://sistemas.minam.gob.pe/SigersolMunicipal/#/panel>

- MINAM. (2021). *Listado de relenos sanitarios*.
<https://www.gob.pe/institucion/minam/informes-publicaciones/279709-listado-de-rellenos-sanitarios>
- Mladenova, R., & Shtereva, D. (2009). Pesticide residues in apples grown under a conventional and integrated pest management system. *Food Additives & Contaminants: Part A*, 26(6), 854–858. <https://doi.org/10.1080/02652030902726060>
- Mohd Yusof, M., Othman, F., Hashim, N., & Ali, N. (2012). The Role of Socio-Economic and Cultural Factors in Municipal Solid Waste Generation: A Case Study in Taman Perling, Johor Bahru. *Jurnal Teknologi, December*. <https://doi.org/10.11113/jt.v37.551>
- Morone, P., Koutinas, A., Gathergood, N., Arshadi, M., & Matharu, A. (2019). Food waste: Challenges and opportunities for enhancing the emerging bio-economy. *Journal of Cleaner Production*, 221, 10–16. <https://doi.org/10.1016/j.jclepro.2019.02.258>
- Mu'azu, N. D., Blaisi, N. I., Naji, A. A., Abdel-Magid, I. M., & AlQahtany, A. (2019). Food waste management current practices and sustainable future approaches: a Saudi Arabian perspectives. *Journal of Material Cycles and Waste Management*, 21(3), 678–690. <https://doi.org/10.1007/s10163-018-0808-4>
- Munda, G. (2004). Social multi-criteria evaluation: Methodological foundations and operational consequences. *European Journal of Operational Research*, 158(3), 662–677. [https://doi.org/10.1016/S0377-2217\(03\)00369-2](https://doi.org/10.1016/S0377-2217(03)00369-2)
- Munda, G. (2014). On the Use of Shadow Prices for Sustainable Well-Being Measurement. *Social Indicators Research*, 118(2), 911–918. <https://doi.org/10.1007/s11205-013-0446-0>
- Munir, M. T., Li, B., & Naqvi, M. (2023). Revolutionizing municipal solid waste management

- (MSWM) with machine learning as a clean resource: Opportunities, challenges and solutions. *Fuel*, 348, 128548. <https://doi.org/10.1016/j.fuel.2023.128548>
- Naciones Unidas. (2021). Construir un futuro mejor. *Cepal*, 1–200.
- Namlis, K. G., & Komilis, D. (2019). Influence of four socioeconomic indices and the impact of economic crisis on solid waste generation in Europe. *Waste Management*, 89, 190–200. <https://doi.org/10.1016/j.wasman.2019.04.012>
- Nguyen, X. C. H., Nguyen, T. T. H., La, D. D., Kumar, G., Rene, E. R., Nguyen, D. D., Chang, S. W., Chung, W. J., Nguyen, X. C. H., & Nguyen, V. K. (2021). Development of machine learning - based models to forecast solid waste generation in residential areas: A case study from Vietnam. *Resources, Conservation and Recycling*, 167(July 2020), 105381. <https://doi.org/10.1016/j.resconrec.2020.105381>
- OCDE. (2015). *Material Resources, Productivity and the Environment*. OECD. <https://doi.org/10.1787/9789264190504-en>
- ONU. (2023). *Informe de los Objetivos de Desarrollo Sostenible 2023: Edición Especial*. https://unstats.un.org/sdgs/report/2023/The-Sustainable-Development-Goals-Report-2023_Spanish.pdf?_gl=1*_arkxzk*_ga*MTc3ODI5OTYwMS4xNjE4MzU3NTMz*_ga_TK9BQL5X7Z*MTY5NDkwMzM3OS4xLjEuMTY5NDkwMzU1OS4wLjAuMA..
- Ordoñez, P. (2004). *A Model for Assessing Waste Generation Factors and Forecasting Waste Generation using Artificial Neural Networks: A Case Study of Chile*.
- Orias, N. E., Reynolds, C. J., Ernstoff, A. S., Vázquez-Rowe, I., Cooper, K., & Aldaco, R. (2021). Editorial: Food Loss and Waste: Not All Food Waste Is Created Equal. *Frontiers in Nutrition*, 8, 2–3. <https://doi.org/10.3389/fnut.2021.615550>
- Orihuela, J. (2018). Un Análisis de la Eficiencia de la Gestión Municipal de Residuos Sólidos

en el Perú y sus determinantes. *Instituto Nacional de Estadística*, 1–82.

OSINERGMIN. (2019). *Energy Investment Supervisory Agency*.

OSINERGMIN. (2020a). *Natural Gas Consumption and Uses Results Report Residential Energy Use and Consumption Survey - ERCUE 2019-2020*.

OSINERGMIN. (2020b). *Results Report on Consumption and Uses of Liquid Hydrocarbons and LPG Residential Energy Consumption and Uses Survey - ERCUE 2019-2020*.

OSINERGMIN. (2021a). *Electricity Consumption and Uses Results Report Residential Energy Consumption and Uses Survey - ERCUE 2019-2020*.
<https://doi.org/https://www.gob.pe/institucion/osinergmin/informes-publicaciones/2507564-encuesta-residencial-de-consumo-y-usos-de-energia-ercue-2019-2020>

OSINERGMIN. (2021b). *Statistical bulletin: natural gas processing, production and transportation*. (p. 39).

Patwa, A., Parde, D., Dohare, D., Vijay, R., & Kumar, R. (2020). Solid waste characterization and treatment technologies in rural areas: An Indian and international review. *Environmental Technology and Innovation*, 20, 101066.
<https://doi.org/10.1016/j.eti.2020.101066>

Paulauskaite-Taraseviciene, A., Raudonis, V., & Sutiene, K. (2022). Forecasting municipal solid waste in Lithuania by incorporating socioeconomic and geographical factors. *Waste Management*, 140(October 2021), 31–39. <https://doi.org/10.1016/j.wasman.2022.01.004>

Paulraj, S. G., Hait, S., & Thakur, A. (2016). Automated municipal solid waste sorting for recycling using a mobile manipulator. *Proceedings of the ASME Design Engineering Technical Conference*, 5A-2016, 1–10. <https://doi.org/10.1115/DETC2016-59842>

- Pirani, S. I., Al-Khatib, I. A., Halaweh, R., Arafat, M. A., & Arafat, H. A. (2015). Household-level determinants of residential solid waste generation rates: a study from Nablus-Palestine. *Journal of Material Cycles and Waste Management*, 17(4), 725–735. <https://doi.org/10.1007/s10163-014-0304-4>
- Popli, K., Park, C., Han, S.-M., & Kim, S. (2021). Prediction of Solid Waste Generation Rates in Urban Region of Laos Using Socio-Demographic and Economic Parameters with a Multi Linear Regression Approach. *Sustainability*, 13(6), 3038. <https://doi.org/10.3390/su13063038>
- Powell, J. T., Chertow, M. R., & Esty, D. C. (2018). Where is global waste management heading? An analysis of solid waste sector commitments from nationally-determined contributions. *Waste Management*, 80, 137–143. <https://doi.org/10.1016/j.wasman.2018.09.008>
- Qu, X., Li, Z., Xie, X., Sui, Y., Yang, L., & Chen, Y. (2009). Survey of composition and generation rate of household wastes in Beijing, China. *Waste Management*, 29(10), 2618–2624. <https://doi.org/10.1016/j.wasman.2009.05.014>
- Rodríguez-Jiménez, L. M., Pérez-Vidal, A., & Torres-Lozada, P. (2022). Research trends and strategies for the improvement of anaerobic digestion of food waste in psychrophilic temperatures conditions. *Heliyon*, 8(10). <https://doi.org/10.1016/j.heliyon.2022.e11174>
- Scarlat, N., Motola, V., Dallemand, J. F., Monforti-Ferrario, F., & Mofor, L. (2015). Evaluation of energy potential of Municipal Solid Waste from African urban areas. *Renewable and Sustainable Energy Reviews*, 50, 1269–1286. <https://doi.org/10.1016/j.rser.2015.05.067>
- Schanes, K., Dobernig, K., & Gözet, B. (2018). Food waste matters - A systematic review of household food waste practices and their policy implications. *Journal of Cleaner Production*, 182, 978–991. <https://doi.org/10.1016/j.jclepro.2018.02.030>

- Serrona, K. R., & soo Yu, J. (2009). Finding urban waste management solutions and policies: Waste-to-energy development and livelihood support system in Payatas, Metro Manila, Philippines. *Journal of Environmental Sciences*, 21(SUPPL. 1), S40–S43. [https://doi.org/10.1016/S1001-0742\(09\)60033-4](https://doi.org/10.1016/S1001-0742(09)60033-4)
- Shah, A. V, Srivastava, V. K., Mohanty, S. S., & Varjani, S. (2021). Municipal solid waste as a sustainable resource for energy production: State-of-the-art review. *Journal of Environmental Chemical Engineering*, 9(4), 105717. <https://doi.org/10.1016/j.jece.2021.105717>
- Sharma, A., Ganguly, R., & Gupta, A. K. (2019). Characterization and Energy Generation Potential of Municipal Solid Waste from Nonengineered Landfill Sites in Himachal Pradesh, India. *Journal of Hazardous, Toxic, and Radioactive Waste*, 23(4), 04019008. [https://doi.org/10.1061/\(ASCE\)HZ.2153-5515.0000442](https://doi.org/10.1061/(ASCE)HZ.2153-5515.0000442)
- Sinha, R., & Prabhudev, B. (2016). Impact of socio cultural challenges in solid waste management. *International Journal of Engineering Research & Technology*, 4(27).
- Springmann, M., Clark, M., Mason-D’Croz, D., Wiebe, K., Bodirsky, B. L., Lassaletta, L., de Vries, W., Vermeulen, S. J., Herrero, M., Carlson, K. M., Jonell, M., Troell, M., DeClerck, F., Gordon, L. J., Zurayk, R., Scarborough, P., Rayner, M., Loken, B., Fanzo, J., ... Willett, W. (2018). Options for keeping the food system within environmental limits. *Nature*, 562(7728), 519–525. <https://doi.org/10.1038/s41586-018-0594-0>
- Sujauddin, M., Huda, S. M. S., & Hoque, A. T. M. R. (2008). Household solid waste characteristics and management in Chittagong, Bangladesh. *Waste Management*, 28(9), 1688–1695. <https://doi.org/10.1016/j.wasman.2007.06.013>
- Sukholthaman, P., Chanvarasuth, P., & Sharp, A. (2017). Analysis of waste generation variables and people’s attitudes towards waste management system: a case of Bangkok,

- Thailand. *Journal of Material Cycles and Waste Management*, 19(2), 645–656.
<https://doi.org/10.1007/s10163-015-0456-x>
- Sunayana, Kumar, S., & Kumar, R. (2021). Forecasting of municipal solid waste generation using non-linear autoregressive (NAR) neural models. *Waste Management*, 121(2021), 206–214. <https://doi.org/10.1016/j.wasman.2020.12.011>
- Suthar, S., & Singh, P. (2015). Household solid waste generation and composition in different family size and socio-economic groups: A case study. *Sustainable Cities and Society*, 14(1), 56–63. <https://doi.org/10.1016/j.scs.2014.07.004>
- Thomson Reuters. (2024). *ISI Web of Knowledge*.
- Tomić, T., & Schneider, D. R. (2017). Municipal solid waste system analysis through energy consumption and return approach. *Journal of Environmental Management*, 203, 973–987.
<https://doi.org/10.1016/j.jenvman.2017.06.070>
- Torrente-Velásquez, J. M., Ripa, M., Chifari, R., & Giampietro, M. (2021). Identification of inference fallacies in solid waste generation estimations of developing countries. A case-study in Panama. *Waste Management*, 126, 454–465.
<https://doi.org/10.1016/j.wasman.2021.03.037>
- Trang, P. T. T., Dong, H. Q., Toan, D. Q., Hanh, N. T. X., & Thu, N. T. (2017). The Effects of Socio-economic Factors on Household Solid Waste Generation and Composition: A Case Study in Thu Dau Mot, Vietnam. *Energy Procedia*, 107(September 2016), 253–258.
<https://doi.org/10.1016/j.egypro.2016.12.144>
- UN-HABITAT. (2010). *Solid Waste Management in the World's Cities – Water and Sanitation in the World's Cities 2010*. <https://doi.org/10.1108/meq.2010.21.6.869.3>
- UNEP. (2021). *Food Waste Index Report 2021* (p. 100).

- Valente de Macedo, L. S., Barda Picavet, M. E., Puppim de Oliveira, J. A., & Shih, W. Y. (2021). Urban green and blue infrastructure: A critical analysis of research on developing countries. *Journal of Cleaner Production*, 313(July 2020). <https://doi.org/10.1016/j.jclepro.2021.127898>
- van Dooren, C., Mensink, F., Eversteijn, K., & Schrijnen, M. (2020). Development and Evaluation of the Eetmaatje Measuring Cup for Rice and Pasta as an Intervention to Reduce Food Waste. *Frontiers in Nutrition*, 6(February), 1–11. <https://doi.org/10.3389/fnut.2019.00197>
- Vázquez-Rowe, I., Kahhat, R., Larrea-Gallegos, G., & Ziegler-Rodriguez, K. (2019). Peru's road to climate action: Are we on the right path? The role of life cycle methods to improve Peruvian national contributions. In *Science of the Total Environment* (Vol. 659, pp. 249–266). Elsevier B.V. <https://doi.org/10.1016/j.scitotenv.2018.12.322>
- Vázquez-Rowe, I., Ziegler-Rodriguez, K., Margallo, M., Kahhat, R., & Aldaco, R. (2021). Climate action and food security: Strategies to reduce GHG emissions from food loss and waste in emerging economies. *Resources, Conservation and Recycling*, 170(October 2020). <https://doi.org/10.1016/j.resconrec.2021.105562>
- Véliz, C. (2023). *Machine learning. Introduction to deep learning*. Fondo Editorial de la PUCP.
- Velusamy, P., Srinivasan, J., Subramanian, N., Mahendran, R. K., Saleem, M. Q., Ahmad, M., Shafiq, M., & Choi, J. G. (2023). Optimization-Driven Machine Learning Approach for the Prediction of Hydrochar Properties from Municipal Solid Waste. *Sustainability (Switzerland)*, 15(7). <https://doi.org/10.3390/su15076088>
- Viktorin, A., Hrabec, D., Nevrlý, V., Šomplák, R., & Šenkeřík, R. (2023). Hierarchical clustering-based algorithms for optimal waste collection point locations in large-scale problems: A framework development and case study. *Computers & Industrial*

Engineering, 178(March), 109142. <https://doi.org/10.1016/j.cie.2023.109142>

- Vitorino de Souza Melaré, A., Montenegro González, S., Faceli, K., & Casadei, V. (2017). Technologies and decision support systems to aid solid-waste management: a systematic review. *Waste Management*, 59, 567–584. <https://doi.org/10.1016/j.wasman.2016.10.045>
- Vu, H. L., Ng, K. T. W., & Bolingbroke, D. (2019). Time-lagged effects of weekly climatic and socio-economic factors on ANN municipal yard waste prediction models. *Waste Management*, 84, 129–140. <https://doi.org/10.1016/j.wasman.2018.11.038>
- Vyas, S., Prajapati, P., Shah, A. V., & Varjani, S. (2022). Municipal solid waste management : Dynamics , risk assessment , ecological influence , advancements , constraints and perspectives. *Science of the Total Environment*, 814, 152802. <https://doi.org/10.1016/j.scitotenv.2021.152802>
- Withanage, S. V., Dias, G. M., & Habib, K. (2021). Review of household food waste quantification methods: Focus on composition analysis. *Journal of Cleaner Production*, 279, 123722. <https://doi.org/10.1016/j.jclepro.2020.123722>
- Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, 39(1), 93–112. <https://doi.org/10.1177/0739456X17723971>
- Xu, A., Chang, H., Xu, Y., Li, R., Li, X., & Zhao, Y. (2021). Applying artificial neural networks (ANNs) to solve solid waste-related issues: A critical review. *Waste Management*, 124, 385–402. <https://doi.org/10.1016/j.wasman.2021.02.029>
- Xu, D. Y., Lin, Z. Y., Gordon, M. P. R., Robinson, N. K. L., & Harder, M. K. (2016). Perceived key elements of a successful residential food waste sorting program in urban apartments: stakeholder views. *Journal of Cleaner Production*, 134(Part A), 362–370.

<https://doi.org/10.1016/j.jclepro.2015.12.107>

Younes, M. K., Nopiah, Z. M., Basri, N. E. A., Basri, H., Abushammala, M. F. M., & Maulud, K. N. A. (2015). Prediction of municipal solid waste generation using nonlinear autoregressive network. *Environmental Monitoring and Assessment*, 187(12), 1–10. <https://doi.org/10.1007/s10661-015-4977-5>

Zaman, A. U. (2014). Identification of key assessment indicators of the zero waste management systems. *Ecological Indicators*, 36, 682–693. <https://doi.org/10.1016/j.ecolind.2013.09.024>

Zhu, J. J., Park, D., Chang, D. T., Cheng, C., Anderson, P. R., & Fan, H. J. (2021). Unsupervised aided investigation on the associations between municipal solid waste characteristics and socio-economic conditions. *Environmental Research*, 194(July 2020), 110633. <https://doi.org/10.1016/j.envres.2020.110633>

Ziegler-Rodriguez, K., Margallo, M., Aldaco, R., Vázquez-Rowe, I., & Kahhat, R. (2019). Transitioning from open dumpsters to landfilling in Peru: Environmental benefits and challenges from a life-cycle perspective. *Journal of Cleaner Production*, 229, 989–1003. <https://doi.org/10.1016/j.jclepro.2019.05.015>

APÉNDICE - RESÚMENES DE PRODUCCIÓN CIENTÍFICA

- **Manuscrito 1 – Artículo de revisión**

“Reviewing the influence of sociocultural, environmental and economic variables to forecast municipal solid waste (MSW) generation”

Izquierdo-Horna, L., Kahhat, R., & Vázquez-Rowe, I. (2022). Reviewing the influence of sociocultural, environmental and economic variables to forecast municipal solid waste (MSW) generation. *Sustainable Production and Consumption*, 33, 809–819.
<https://doi.org/10.1016/j.spc.2022.08.008>

Abstract:

Municipal solid waste (MSW) generation forecasting has become an important tool for decision-making in urban environments, not only due to its essential role in effective waste management, but also because it provides an understanding of the complexity of the factors that govern it. Current research bases its forecast models (e.g., artificial neural networks, regression methods, three decision methods...) on predictive variables supported by pre-existing government information or, alternatively, on related studies with different site characteristics due to the lack of primary data from the specific sector. These assumptions and generalizations generate a different representation of the area of interest, raising the level of uncertainty of the results and reducing their level of reliability. Thus, the current review focuses on exploring the influence, relevance and opportunities for improvement when it comes to including or excluding sociocultural, environmental and/or economic variables in the solid waste forecasting process. Relevant information has been provided regarding the predictor variables considered to have better predictive power and, at the same time, limitations in data availability have been highlighted. Finally, it is concluded that the adoption of case study-specific predictor variables collected through primary data (e.g., questionnaires or surveys)

would improve the predictive performance of the models providing a robust and effective tool for waste management. In addition, it is expected that the recommendations provided will be useful for future research related to the prediction of MSW and, thus, contribute to obtaining more representative results.

Keywords: Forecasting, Influencing variables, Semi-systematic review, Solid waste generation, Waste management

- **Manuscrito 2 – Ponencia de congreso**

“Identification of variables related to household solid waste generation forecasting in residential areas”

Izquierdo-Horna, L., Kahhat, R., & Vázquez-Rowe, I. (2023). Identification of variables related to household solid waste generation forecasting in residential areas. 19th International Symposium on Waste Management, Resource Recovery and Sustainable Landfilling

Abstract:

Household solid waste generation has become one of the main environmental challenges in recent years. Several strategies and techniques have been proposed to contribute to an adequate waste management, where forecasting has been the most popular and accepted due to the nature of the available data. In this sense, the main objective of this research is to determine a set of variables potentially capable of segmenting and characterizing certain contexts with respect to household solid waste generation. For this purpose, different open-access governmental databases were reviewed and the information was systematized to be able to catalog the variables identified in socioeconomic, socio-cultural and environmental dimensions. Preliminary results show that the variables associated with monthly income, education, age, population density and energy consumption are the most relevant. This analysis will help decision-makers to optimize their prediction models and obtain better results.

Keywords: Economic variables, Environmental variables, Machine Learning, Sociocultural variables, Waste management.

- **Manuscrito 3 – Artículo original**

“Unveiling the energy consumption-food waste nexus in households: A focus on key predictors of food waste generation”

Izquierdo-Horna, L., Kahhat, R. & Vázquez-Rowe, I. (2024). Unveiling the energy consumption-food waste nexus in households: A focus on key predictors of food waste generation. *J Mater Cycles Waste Manag.* <https://doi.org/10.1007/s10163-024-01946-2>

Abstract

In most cities worldwide, household food waste constitutes a significant portion of municipal solid waste (MSW). However, its management often proves inadequate due to the insufficient resources allocated to waste management systems, the omission of the resource potential in MSW, and the lack of recognition of household food waste drivers for forecasting generation in specific geographical contexts. This research aims to identify social, economic, and environmental variables serving as proxies to forecast household food waste generation. To achieve this, a multiple linear regression model was proposed to assess the relationship between cooking fuel type (i.e., liquefied petroleum gas, natural gas, and electricity), land use categories (i.e., commercial, industrial, and residential), population density, expenditure on in-house food consumption, and household food waste generation. Three alternate modeling scenarios were considered based on available data, with Lima, Peru, serving as a case study. The results indicate that the combined consumption of liquefied petroleum gas and natural gas, and electricity consumption, along with residential land use, were the most influential variables. Finally, for a comprehensive understanding of the studied phenomenon, it is crucial to analyze and consider the intricate dynamics of societal consumption patterns.

Keywords: Food waste, Fuel use, Household consumption, Land use, Industrial ecology, Municipal solid waste.

- **Manuscrito 4 – Artículo original**

“Forecasting municipal solid waste generation from household fuel consumption utilizing random forests” – In preparation.



APÉNDICE – BASES DE DATOS UTILIZADAS

La información utilizada en este proyecto está destinada a ser accesible para revisiones anónimas, comentarios y con el objetivo de facilitar la reproducibilidad de los resultados obtenidos. Estamos comprometidos con la transparencia y la colaboración en la investigación científica; por ello, a continuación, se detallan las bases de datos utilizadas.

- **Encuesta Residencial de Consumo y Usos de Energía - ERCUE:**

Información relacionada al consumo de combustible, la cual permite evaluar los patrones de consumo y brechas de acceso de diversas fuentes de energía a nivel residencial en el Perú, así como su relación con las condiciones socioeconómicas de los hogares.

- Link de descarga (acceso abierto):

<https://www.gob.pe/institucion/osinergmin/informes-publicaciones/2507564-encuesta-residencial-de-consumo-y-usos-de-energia-ercue-2019-2020>

Tabla A 1 - Diccionario de variables utilizadas – Sección Hogares – ERCUE 2019.

Variable	Valor	Etiqueta
DEPART	Departamento	
PROVIN	Provincia	
DISTRI	Distrito	
	¿La vivienda que ocupa es...?	
	1	Casa independiente
Y1	2	Departamento en edificio
	3	Vivienda en quinta
	4	Vivienda en Casa de Vecindad (Callejón, Solar o Corralón)

	5	Choza o Cabaña
	6	Vivienda Improvisada (Estera, Caña Chancada, etc.)
	7	Local No Destinado para la Habitación Humana
	8	Otro
	Cocinar sus alimentos (Incluye Hervir el Agua):Electricidad	
Y7B_A_1	1	Fuente principal
	2	Fuente secundaria
	3	Fuente exclusiva
	Cocinar sus alimentos (Incluye Hervir el Agua):Gas GLP	
Y7B_A_2	1	Fuente principal
	2	Fuente secundaria
	3	Fuente exclusiva
	Cocinar sus alimentos (Incluye Hervir el Agua):Gas Natural	
Y7B_A_3	1	Fuente principal
	2	Fuente secundaria
	3	Fuente exclusiva
Y23_5F	Pago Mensual S/. (último mes) Electricidad	
	Dentro de su hogar cuenta con acceso a servicio de: - Gas Natural	
Y23_1G	0	No
	1	Sí
Y23_5G	Pago Mensual S/. (último mes) Gas Natural	
	¿Dentro de su hogar cuenta con el servicio de GLP?	
Y23_1H	0	No
	1	Sí

Y23_5H	Pago mensual S/ (último mes) GLP	
	1 ¿Qué Medio utiliza para Abastecerse de GLP en su Hogar?	
X2_1A	1	GLP en balón
	2	GLP a granel
	¿En su hogar consume balones de GLP de 5kg?	
X2_2A_5k	0	No
	1	Sí
	¿En su hogar consume balones de GLP de 10kg?	
X2_2A_10k	0	No
	1	Sí
	¿En su hogar consume balones de GLP de 15kg?	
X2_2A_15k	0	No
	1	Sí
	¿En su hogar consume balones de GLP de 45kg?	
X2_2A_45k	0	No
	1	Sí
X2_2B_5k	¿Cuál fue el precio de la última adquisición? 5kg	
X2_2B_10k	¿Cuál fue el precio de la última adquisición? 10kg	
X2_2B_15k	¿Cuál fue el precio de la última adquisición? 15kg	
X2_2B_45k	¿Cuál fue el precio de la última adquisición? 45kg	
	¿Con qué frecuencia realiza este consumo? 5kg	
X2_2C_5k	1	Semanal
	2	Quincenal
	3	Mensual

	4	Bimestral
	5	Trimestral
	6	Cuatrimestral
	7	Semestral
	8	Anual

¿Con qué frecuencia realiza este consumo? 10kg

X2_2C_10k	1	Semanal
	2	Quincenal
	3	Mensual
	4	Bimestral
	5	Trimestral
	6	Cuatrimestral
	7	Semestral
	8	Anual

¿Con qué frecuencia realiza este consumo? 15kg

X2_2C_15k	1	Semanal
	2	Quincenal
	3	Mensual
	4	Bimestral
	5	Trimestral
	6	Cuatrimestral
	7	Semestral
	8	Anual

X2_2C_45k ¿Con qué frecuencia realiza este consumo? 45kg

1	Semanal
2	Quincenal
3	Mensual
4	Bimestral
5	Trimestral
6	Cuatrimstral
7	Semestral
8	Anual
ubigeo	Código de ubigeo
gasto_total	Gasto anual total (S/)
n_pers	Número de miembros del hogar
conskwprom	Consumo promedio mensual por hogar (kWh-mes)

Tabla A 2 - Diccionario de variables utilizadas – Sección Personas – ERCUE 2019.

Variable	Valor	Etiqueta
		¿Cuál es la relación de parentesco con el Jefe del Hogar?
	1	Jefe/Jefa
	2	Esposa (o)
	3	Hijo (a)
F1	4	Yerno/Nuera
	5	Padres/Suegros
	6	Otros parientes
	7	Trabajador del Hogar
	8	Pensionista

	9	Otros no parientes
F3	¿Cuál es su edad en años cumplidos?	
	¿Cuál fue el último nivel de estudios alcanzado?	
	1	Sin nivel
	2	Educación inicial
	3	Primaria Incompleta
	4	Primaria Completa
	5	Secundaria Incompleta
F5	6	Secundaria Completa
	7	Superior no Universitaria Incompleta
	8	Superior no Universitaria Completa
	9	Superior Universitaria Incompleta
	10	Superior Universitaria Completa
	11	Post Grado Universitario

- Sistema de información para la gestión de residuos sólidos - SIGERSOL:**
 Plataforma administrada por el Ministerio del Ambiente (MINAM), la cual facilita el registro, procesamiento y difusión de la información sobre la gestión y manejo de residuos sólidos de las municipalidades. Esta herramienta permite conocer los datos asociados a la generación de residuos sólidos en función de los reportes municipales. Asimismo, también facilita información sobre los residuos sólidos alimenticios domiciliarios generados durante el año evaluado dentro del total caracterizado como residuo orgánico.

- Link de descarga (acceso abierto):

<https://sistemas.minam.gob.pe/SigersolMunicipal/#!/accesoLibre/generacion>

Tabla A 3 - Diccionario de variables utilizadas – Sección Generación de RSM - SIGERSOL 2019.

Variable	Valor
DEPARTAMENTO	Departamento
ROVINCIA	Provincia
DISTRITO	Distrito
AÑO DEL REPORTE	Año de referencia - 2019
RsSs Domiciliarios	Residuos sólidos domiciliarios
RsSs No Domiciliarios	Residuos sólidos no domiciliarios
Percápita RsSs Dom.	Generación percápita de los residuos sólidos domiciliarios
Composición RsSs Dom.	Composición de residuos sólidos domiciliarios

Resumen de Reportes Anuales – SIGERSOL

Sección destinada a los resúmenes generales consolidados por cada distrito. Asimismo, también facilita información sobre los residuos sólidos alimenticios generados durante el año en función del porcentaje dentro del total caracterizado como residuo orgánico.

- Link de descarga (acceso abierto):

<https://sistemas.minam.gob.pe/SigersolMunicipal/#!/accesoLibre/resumenes>

Por otro lado, de manera complementaria, la Oficina de Gestión Documental y Atención a la Ciudadanía – MINAM, a través del canal Atención Ciudadanía, provee bases integradas de SIGERSOL – Municipal para el año 2019. Esta base integrada, es el resultado del procesamiento que MINAM hace a los datos registrados directamente en SIGERSOL.

- **Base de datos utilizada para relacionar la generación de residuos alimenticios domésticos y el consumo de combustible domiciliario**

Tabla A 4 - Base de datos utilizadas para relacionar el consumo de combustible doméstico y la generación per cápita de los residuos sólidos alimenticios en hogares – Variable independiente: Uso de suelo.

Distrito	Variable dependiente	Variable independiente – Uso de suelo		
	GPC_FW	Suelo residencial	Suelo industrial	Suelo comercial
Ancón	0.383	0.037	0.008	0.001
Ate Vitarte	0.285	0.567	0.124	0.065
Barranco	0.26	0.663	0	0.13
Bellavista	0.333	0.691	0.003	0.08
Breña	0.39	0.703	0.026	0.105
Callao	0.362	0.523	0.016	0.021
Carabaylo	0.346	0.472	0.043	0.02
Carmen de la Legua Reynoso	0.438	0.574	0.008	0.008
Chaclacayo	0.301	0.657	0.114	0.039
Chorrillos	0.301	0.461	0.058	0.034
Cieneguilla	0.28	0.533	0.013	0.016
Comas	0.443	0.615	0.081	0.057
El Agustino	0.448	0.513	0.098	0.036
Independencia	0.297	0.702	0.066	0.102
Jesús María	0.242	0.535	0	0.089
La Molina	0.349	0.621	0	0.029
La Perla	0.386	0.434	0.012	0.043
La Punta	0.329	0.506	0.019	0.011
La Victoria	0.304	0.484	0.01	0.317
Lima	0.394	0.331	0.2	0.189
Lince	0.282	0.647	0.002	0.195
Los Olivos	0.388	0.624	0.072	0.089
Lurigancho	0.285	0.42	0.165	0.022
Lurín	0.359	0.135	0.271	0.067
Magdalena del Mar	0.323	0.69	0	0.046
Mi Perú	0.256	0.023	0	0
Miraflores	0.348	0.779	0	0.048
Pachacamac	0.288	0.367	0.012	0.016

Pucusana	0.289	0.279	0.006	0.017
Pueblo Libre	0.34	0.644	0.015	0.076
Puente Piedra	0.336	0.515	0.167	0.062
Punta Hermosa	0.311	0.093	0.119	0.003
Punta Negra	0.254	0.02	0	0
Rímac	NA	0.568	0.012	0.055
San Bartolo	0.262	0.493	0	0.035
San Borja	0.306	0.64	0	0.054
San Isidro	0.463	0.137	0	0.286
San Juan de Lurigancho	0.185	0.701	0.028	0.038
San Juan de Miraflores	0.35	0.629	0.021	0.045
San Luis	0.393	0.453	0.147	0.173
San Martín de Porres	0.278	0.725	0.025	0.071
San Miguel	0.366	0.53	0.006	0.088
Santa Anita	0.287	0.46	0.121	0.125
Santa María del Mar	NA	0.36	0	0.015
Santa Rosa	NA	0.584	0.004	0
Santiago de Surco	0.318	0.6	0.004	0.047
Surquillo	0.275	0.571	0.023	0.22
Ventanilla	0.35	0.001	0.054	0
Villa el Salvador	0.355	0.435	0.137	0.099
Villa María del Triunfo	0.445	0.645	0.033	0.03

Tabla A 5 - Base de datos utilizadas para relacionar el consumo de combustible doméstico y la generación per cápita de los residuos sólidos alimenticios en hogares – Variable independiente: Consumo de combustibles para cocinar en el hogar.

Distrito	Variable independiente – Consumo de combustible para cocinar					
	Ratio GLP	Pago GLP	Ratio electricidad	Pago electricidad	Ratio GN	Pago GN
Ancón	0.947	469.0	0.025	673.1	NA	NA
Ate Vitarte	0.845	397.3	0.029	956.0	0.103	296.8
Barranco	0.881	489.6	0.126	1223.4	NA	NA
Bellavista	0.95	452.8	0.071	1328.6	NA	NA
Breña	0.903	357.0	0.061	1132.4	0.019	NA
Callao	0.765	404.2	0.049	1029.2	0.192	297.4
Carabayllo	0.918	415.6	0.029	765.1	0.04	321.6

Carmen de la Legua Reynoso	0.947	353.4	0.034	1104.0	NA	205.2
Chaclacayo	0.959	430.8	0.047	1039.2	NA	NA
Chorrillos	0.951	459.0	0.044	1315.0	NA	420.0
Cieneguilla	0.933	522.0	0.032	600.0	NA	NA
Comas	0.611	437.2	0.032	1072.3	0.359	292.3
El Agustino	0.644	431.6	0.038	1117.2	0.314	270.0
Independencia	0.741	412.6	0.027	991.2	0.22	250.3
Jesús María	0.881	428.0	0.103	1573.6	0.036	180.0
La Molina	0.905	465.5	0.134	1938.8	NA	432.0
La Perla	0.95	404.2	0.067	1137.4	NA	NA
La Punta	0.882	480.0	0.137	1440.0	NA	NA
La Victoria	0.913	394.7	0.054	1027.7	NA	456.0
Lima	0.835	407.8	0.052	1089.2	0.093	287.6
Lince	0.879	404.6	0.109	1380.6	0.014	NA
Los Olivos	0.573	392.0	0.06	1100.4	0.389	288.4
Lurigancho	0.944	404.5	0.025	767.0	NA	NA
Lurín	0.858	479.4	0.034	1060.8	0.094	489.0
Magdalena del Mar	0.807	429.8	0.095	1481.3	0.118	525.0
Mi Perú	0.969	387.1	0.03	826.0	NA	292.3
Miraflores	0.811	442.7	0.223	2043.2	NA	NA
Pachacamac	0.94	495.5	0.022	898.0	0.009	NA
Pucusana	0.948	456.0	0.015	1278.0	NA	NA
Pueblo Libre	0.762	395.6	0.099	1697.4	0.169	276.0
Puente Piedra	0.916	427.0	0.028	826.7	0.035	302.6
Punta Hermosa	0.91	376.8	0.045	1200.0	NA	NA
Punta Negra	0.937	NA	0.046	NA	NA	NA
Rímac	0.943	411.4	0.043	1040.8	NA	NA
San Bartolo	0.952	NA	0.045	NA	NA	NA
San Borja	0.886	451.2	0.15	1993.3	NA	492.0
San Isidro	0.704	569.3	0.293	2459.2	0.059	NA
San Juan de Lurigancho	0.809	430.9	0.031	949.4	0.145	323.5
San Juan de Miraflores	0.598	436.3	0.043	1180.7	0.374	319.2
San Luis	0.923	412.7	0.074	1044.0	NA	NA
San Martín de Porres	0.674	439.1	0.042	1109.9	0.286	282.4
San Miguel	0.724	406.9	0.105	1535.0	0.207	379.0

Santa Anita	0.719	414.4	0.035	1066.7	0.223	410.3
Santa María del Mar	0.883	NA	0.087	NA	NA	NA
Santa Rosa	0.936	486.0	0.042	680.0	NA	NA
Santiago de Surco	0.851	441.5	0.117	1675.9	0.057	594.8
Surquillo	0.894	423.6	0.091	1244.8	0.018	NA
Ventanilla	0.955	400.9	0.033	738.8	NA	240.5
Villa el Salvador	0.484	424.7	0.023	878.0	0.498	346.2
Villa María del Triunfo	0.67	407.9	0.028	894.2	0.296	341.5

Tabla A 6 - Base de datos utilizadas para relacionar el consumo de combustible doméstico y la generación per cápita de los residuos sólidos alimenticios en hogares – Variable independiente: Gasto en alimentos.

Distrito	Variable dependiente		Variable independiente
	GPC_FW		Gasto promedio en alimentación por mes
Ancón	0.383		1017.0
Ate Vitarte	0.285		1662.1
Barranco	0.26		727.1
Bellavista	0.333		1163.0
Breña	0.39		691.5
Callao	0.362		868.4
Carabaylo	0.346		978.8
Carmen de la Legua Reynoso	0.438		646.7
Chaclacayo	0.301		758.2
Chorrillos	0.301		1138.5
Cieneguilla	0.28		655.8
Comas	0.443		682.0
El Agustino	0.448		629.9
Independencia	0.297		711.7
Jesús María	0.242		733.0
La Molina	0.349		875.0
La Perla	0.386		698.5
La Punta	0.329		631.3
La Victoria	0.304		768.1
Lima	0.394		1078.0
Lince	0.282		673.5

Los Olivos	0.388	799.2
Lurigancho	0.285	684.9
Lurín	0.359	738.3
Magdalena del Mar	0.323	710.2
Mi Perú	0.256	604.7
Miraflores	0.348	1935.9
Pachacamac	0.288	739.1
Pucusana	0.289	866.2
Pueblo Libre	0.34	1295.2
Puente Piedra	0.336	697.1
Punta Hermosa	0.311	655.4
Punta Negra	0.254	457.1
Rímac	NA	896.0
San Bartolo	0.262	444.4
San Borja	0.306	870.8
San Isidro	0.463	1129.1
San Juan de Lurigancho	0.185	1354.6
San Juan de Miraflores	0.35	1203.0
San Luis	0.393	865.1
San Martín de Porres	0.278	935.9
San Miguel	0.366	719.0
Santa Anita	0.287	626.2
Santa María del Mar	NA	419.0
Santa Rosa	NA	480.0
Santiago de Surco	0.318	995.0
Surquillo	0.275	1126.3
Ventanilla	0.35	1572.6
Villa el Salvador	0.355	1019.0
Villa María del Triunfo	0.445	1318.6

Nota: Toda la información mostrada en esta tabla se basa en el año 2019, donde: GPC_FW: generación de residuos alimentarios domésticos per cápita - Suelo residencial: tasa de uso de suelo residencial por distrito - Suelo industrial: tasa de uso de suelo industrial por distrito -

Suelo comercial: tasa de uso de suelo comercial por distrito - Gasto en alimentación: gasto medio mensual en consumo de alimentos - Ratio GLP: tasa de uso de los hogares que utilizan este combustible para cocinar alimentos – Pago GLP: gasto mensual medio en este combustible - Ratio de electricidad: tasa de uso de los hogares que utilizan este combustible para cocinar alimentos - Pago electricidad: gasto mensual medio en este combustible – Ratio GN: tasa de uso de los hogares que utilizan este combustible (gas natural) para cocinar alimentos – Pago GN: gasto mensual medio en este combustible. NA: No se dispone de información



APÉNDICE – FIGURAS COMPLEMENTARIAS

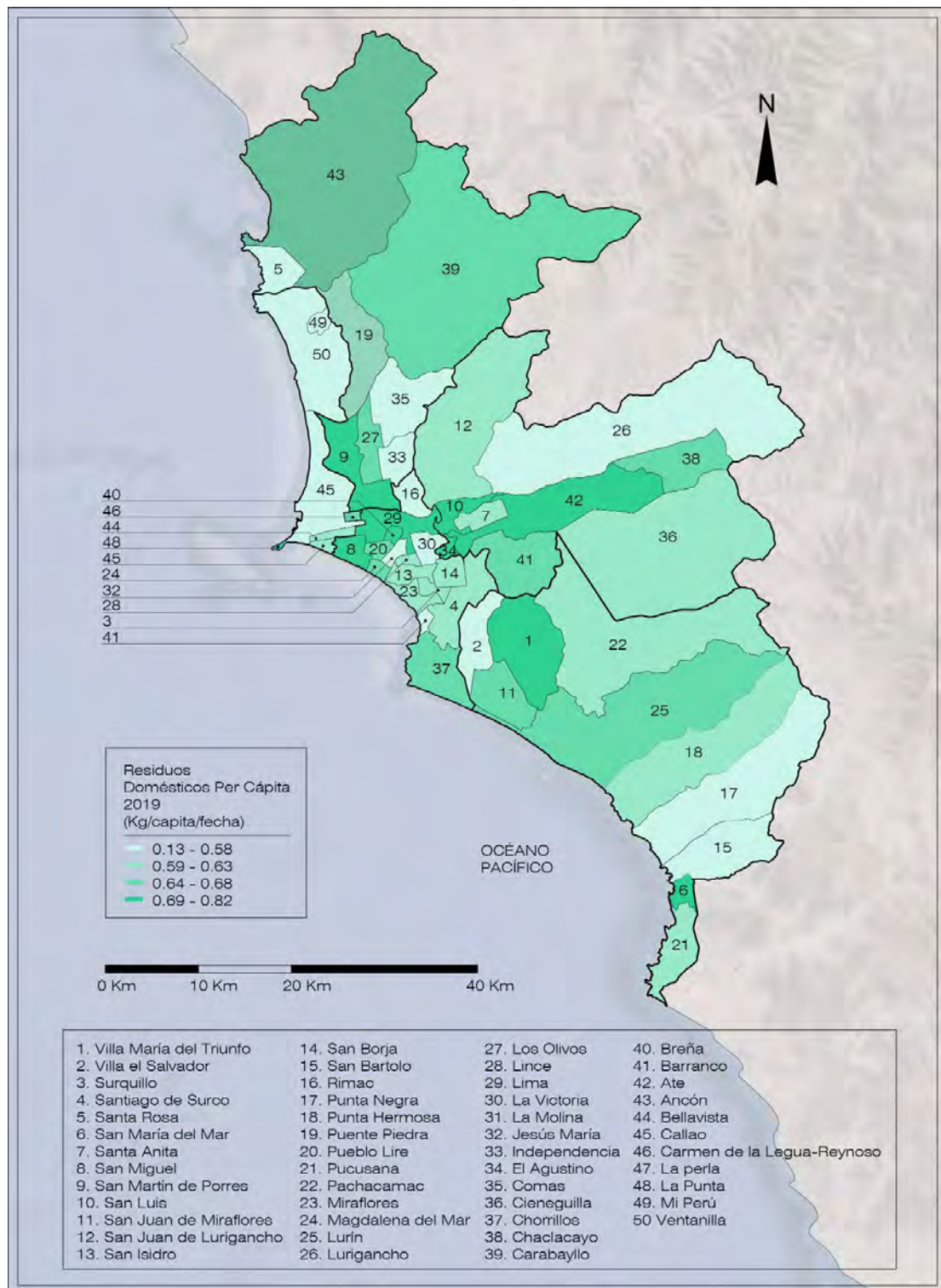


Figura A 1 - Distribución espacial de los residuos sólidos domiciliarios per cápita generados a nivel distrital para Lima Metropolitana en 2019.

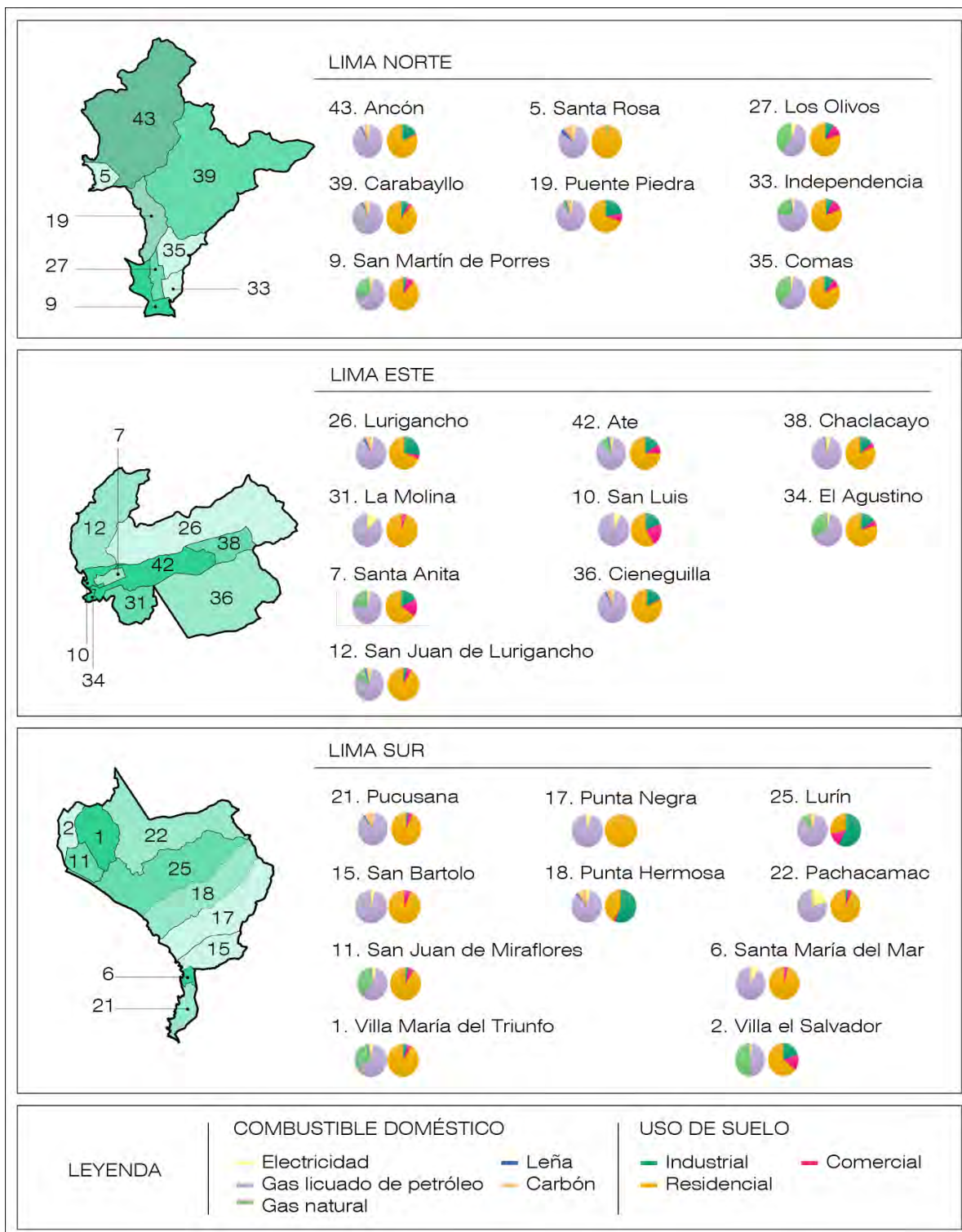


Figura A 2 - Distribución del consumo doméstico de combustible y uso de suelo a nivel distrital para Lima Norte, Lima Este y Lima Sur.

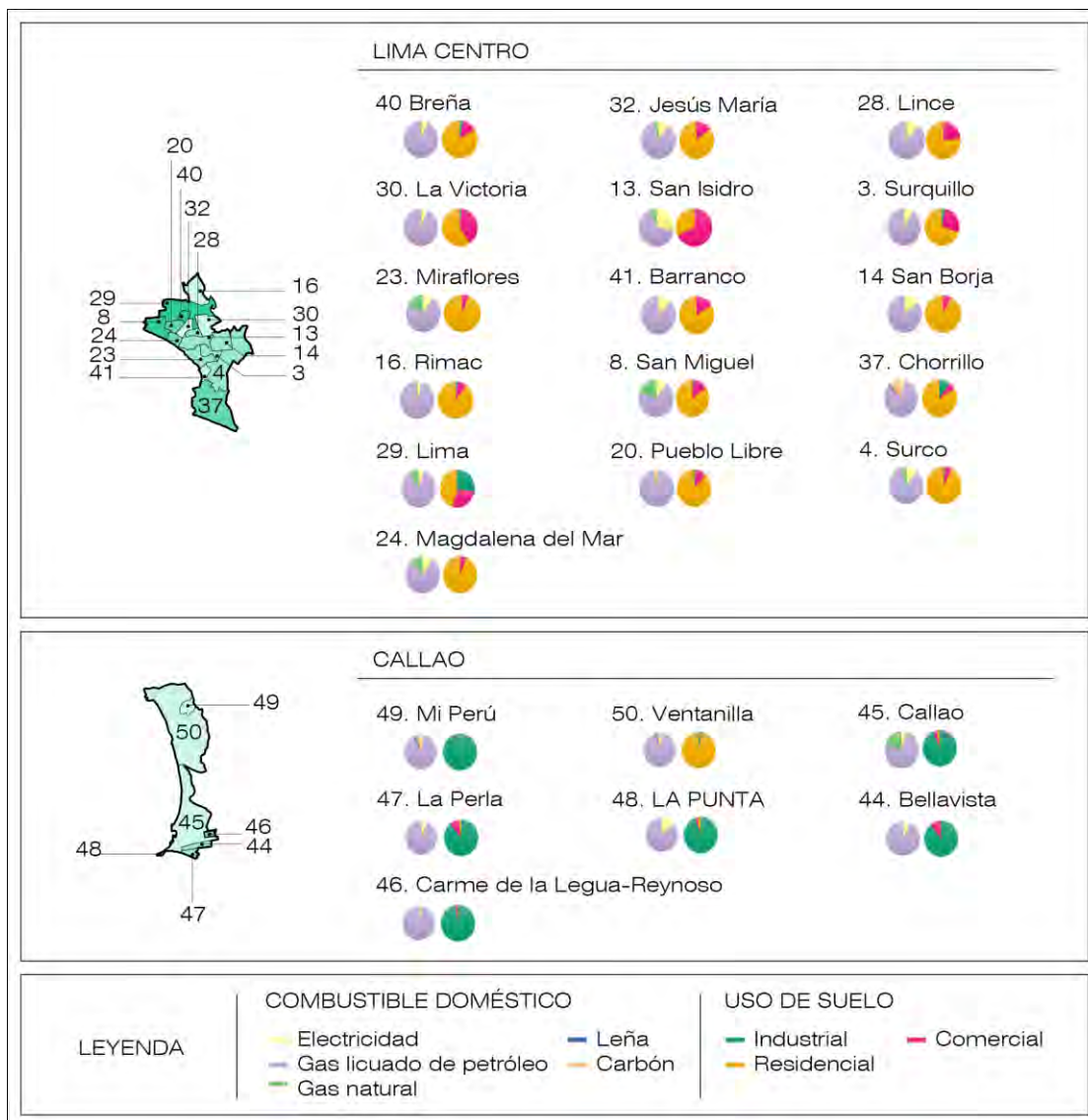


Figura A 3 - Distribución del consumo doméstico de combustible y uso de suelo a nivel distrital para Lima Centro y Callao.

APÉNDICE – EVENTOS DE DIVULGACIÓN

Izquierdo-Horna, L., Kahhat, R., & Vázquez-Rowe, I. (2022). Connection between food waste generation and household fuel use: Trends and implications in Peru. 13th International Conference on Life Cycle Assessment of Food (LCA Foods 2022). Lima, Peru.

Izquierdo-Horna, L., Kahhat, R., & Vázquez-Rowe, I. (2023). Identification of variables related to household solid waste generation forecasting in residential areas. 19th International Symposium on Waste Management, Resource Recovery and Sustainable Landfilling (SARDINIA 2023). Sardinia, Italy.

