



Técnica de ciencia de datos para el pronóstico de consumo de gas natural en la industria siderúrgica

Data science technique for forecasting natural gas consumption in the steel industry

Jesús Abraham Castorena Peña

Facultad de Ingeniería Mecánica y Eléctrica, Universidad Autónoma de Coahuila, Monclova Coahuila, México

jesuscastorenapena@uadec.edu.mx

ORCID: 0000-0002-8833-1159

Alma Jovita Domínguez Lugo

Facultad de Ingeniería Mecánica y Eléctrica, Universidad Autónoma de Coahuila, Monclova Coahuila, México

almadominguez@uadec.edu.mx

ORCID: 0000-0003-4988-4911

José Roberto Cantú González

Escuela de Sistemas, Universidad Autónoma de Coahuila, Unidad Norte, Ciudad Acuña 26235, Coahuila, México

roberto.cantu@uadec.edu.mx

ORCID: 0000-0001-5616-2947

Dalia Margarita Alba Cisneros

Facultad de Ingeniería Mecánica y Eléctrica, Universidad Autónoma de Coahuila, Monclova Coahuila, México

daliaalba@uadec.edu.mx

ORCID: 0009-0009-1033-8244

doi: <https://doi.org/10.36825/RITI.12.26.007>

Recibido: Septiembre 12, 2024

Aceptado: Octubre 25, 2024

Resumen: La programación eficiente de recursos energéticos de gas natural dentro de la industria siderúrgica se encuentra ante muchos desafíos, debido a que este sector no cuenta con las medidas y herramientas necesarias que puedan apoyar su manejo eficiente. En este estudio, se propone un enfoque de pronóstico de consumo de gas natural mediante la técnica predictiva de regresión lineal múltiple. Para el desarrollo del modelo propuesto, se establecen las principales variables relacionadas con el consumo de gas natural que conformarán el modelo predictivo. La evaluación del modelo se llevó a cabo utilizando datos procedentes de una empresa acerera de México. Los resultados del error porcentual absoluto medio, la raíz del error cuadrático medio y la desviación

absoluta media de los modelos de RLM (MAPE: 10.23%, RMSE: 20492.32, DAM: 20009.03) vs. método tradicional (MAPE: 34.05%, RMSE: 93055.92, DAM: 65170.91), reflejan que los modelos propuestos mejoran significativamente la gestión de recursos de gas natural, proporcionando una mejora sustancial en las estimaciones de volumen de consumo de gas natural para una planificación adecuada de su programación energética.

Palabras clave: Ciencia de Datos, Regresión Lineal Múltiple, Pronóstico, Gas Natural, Industria Siderúrgica.

Abstract: Efficient scheduling of natural gas energy resources within the steel industry faces many challenges, as this sector does not have the necessary measures and tools to support its efficient management. In this study, a forecasting approach to natural gas consumption is proposed using the predictive technique of multiple linear regression. For the development of the proposed model, the main variables related to natural gas consumption that will make up the predictive model are established. The evaluation of the model was carried out using data from a steel company in Mexico. The results of mean absolute percentage error, root mean square error and mean absolute deviation of RLM models (MAPE: 10.23%, RMSE: 20492.32, DAM: 20009.03) vs. traditional method (MAPE: 34.05%, RMSE: 93055.92, DAM: 65170.91), reflect that the proposed models significantly improve natural gas resource management, providing substantial improvement in natural gas consumption volume estimates for proper energy scheduling.

Keywords: Data Science, Multiple Linear Regression, Forecasting, Natural Gas, Steel Industry.

1. Introducción

Las organizaciones a nivel mundial se enfrentan cada vez más a un entorno empresarial dinámico y complejo, para lo cual, deben de ser capaces de adaptarse de forma rápida y oportuna ante tales cambios [1]. Por este motivo, es un desafío el gestionar los recursos energéticos dentro de la industria siderúrgica, debido a que este sector es uno de los principales consumidores de energía en el mundo y no cuenta con las medidas y herramientas necesarias que puedan apoyar su manejo eficiente [2], [3], [4].

Diferentes investigaciones han abordado el problema de la administración de recursos energéticos desde diversas perspectivas. Por una parte, proponen la utilización de recursos secundarios como gas de carbón, calor residual y los gases derivados de los procesos siderúrgicos como alternativas de fuentes de energía, con el fin de mejorar la eficiencia energética [5]. Asimismo, también otros autores sostienen que mediante la implementación de una serie de políticas heterogéneas enfocadas a la eficiencia energética podrían optimizar el consumo de energéticos [6].

Estudios recientes han demostrado que, con el rápido desarrollo de la información global, las acereras modernas han tratado de equiparse con sistemas de planificación de recursos empresariales (ERP) y sistema de ejecución de la manufactura (MES) para dar soporte a la buena gestión de recursos energéticos, sin embargo, estas herramientas cumplen únicamente con los requisitos básicos de la transmisión rápida y oportuna de información de flujo de materiales utilizados en las empresas [7], dejando a un lado funcionalidades que den soporte a la planificación, programación, monitoreo, control y proyección de consumo sobre las diferentes fuentes de energía utilizadas en los procesos siderúrgicos actuales [8], ocasionando con ello una ineficiente gestión de recursos energéticos.

El mejorar la buena gestión y administración energética en el sector siderúrgico, también ha sido abordada bajo el enfoque del uso de técnicas predictivas como herramientas de estimación del consumo de energía, con las cuales, puedan establecer de forma más precisa la cantidad de electricidad utilizada por los distintos procesos industriales. Al comprender y establecer la predicción del consumo de energía, las empresas siderúrgicas tienen a su disposición la oportunidad de mejorar sus estrategias de inversión y compra de energía, disminuir los costos de consumo energético e incrementar la productividad de su producción [9].

En México el sector siderúrgico, ocupó el primer lugar de consumo de energía con un 14.3% del consumo industrial final, siendo el gas natural la principal fuente de energía utilizada [10]. Para las empresas acereras, la buena gestión de recursos energéticos es clave, no solo para operar de manera eficiente cada uno de los procesos industriales, sino para lograr una mayor optimización en el uso de ellos, la cual puede traducirse en ahorros monetarios, productividad, aprovechamiento al máximo en los tiempos de producción y soporte en la toma de decisiones para su consumo [11].

Al no contar con las suficientes herramientas o sistemas de gestión de recursos energéticos, la industria siderúrgica se enfrenta constantemente a problemas de cumplimiento de objetivos, estimaciones imprecisas sobre consumo de gas natural, pérdidas económicas y disminución de eficiencia en sus procesos, convirtiéndose así, en una de las principales áreas de interés a mejorar [12]. Por tal motivo la presente investigación tiene como objetivo la propuesta de un modelo matemático basado en regresión lineal múltiple, que sirva como herramienta en el proceso de asignación y selección de gas natural, para mejorar la eficiencia de los recursos energéticos en la industria siderúrgica en México. Asimismo, los resultados ayudarán a eficientizar la planificación, productividad y supervisión de los procesos hacia situaciones actuales.

2. Estado del arte

2.1. Panorama actual de la gestión de recursos energéticos en el sector siderúrgico

La gestión de recursos energéticos hasta hace unas décadas no era un elemento crucial del esquema general de la administración de la industria acerera, hasta que el consumo de recursos no controlado, comenzó a afectar los costos de operación dentro de las compañías. Actualmente, el sector siderúrgico representa una de las industrias con mayor consumo de recursos energéticos, provocando con ello, que este sector se encuentre en una búsqueda constante de herramientas tecnológicas, que apoyen una buena gestión para lograr ahorros significativos de consumo [13], [14], [15]. Dentro de los principales objetivos de la gestión energética se encuentran: la utilización adecuada de los recursos, la planificación y operación de la producción de energía y sus unidades de consumo [16]. Para aprovechar al máximo los recursos energéticos dentro de la industria siderúrgica, los operadores de programación energética tienen que ajustar en tiempo real y oportuno las cantidades de consumo de sus diversos procesos, para equilibrar y mejorar sus planes de operación [17].

Dentro de las principales fuentes de energías más utilizadas se encuentran el gas natural, el coque de carbón y la electricidad. El gas natural es uno de los combustibles más utilizados como fuente de combustión para uso doméstico y uso industrial, particularmente en las industrias como la siderúrgica, la metalúrgica, la petroquímica, etc. Entre el año 2005 y los primeros nueve meses del 2021 el volumen requerido de gas natural creció en un 62.4% pasando de un consumo de 5 mil 89 millones de pies cúbicos diarios (MMpcd) a 8 mil 625 MMpcd, siendo tres sectores los principales consumidores de este insumo, el cual estuvo dividido de la siguiente forma (Figura 1): el sector eléctrico con un 64.7%; seguido del petrolero que representó el 22.2%; y, en tercer lugar, el sector industrial con el 11.5% del total del gas natural demandado [18]. El resto del consumo de gas natural estuvo dividido por los sectores: el residencial (1.0%), los servicios (0.5%) y el autotransporte (0.1%).

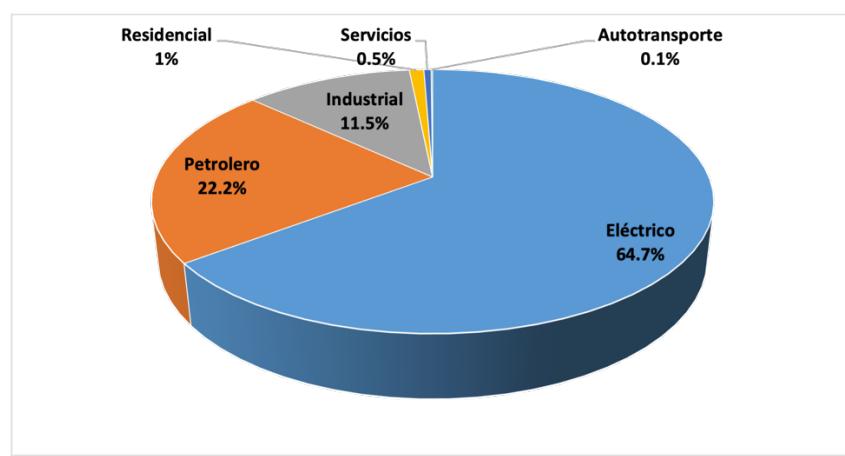


Figura 1. Demanda nacional de gas natural por sector 2020.

Para realizar la predicción del consumo de gas natural dentro del sector siderúrgico, se han utilizado diversos métodos convencionales que van desde la experiencia de los operarios hasta modelos causales, siendo este último método, aplicado a la gestión de gas coque, el cual ha sido desarrollado bajo el enfoque de programación lineal (Figura 2), basado en modelos causales [19].

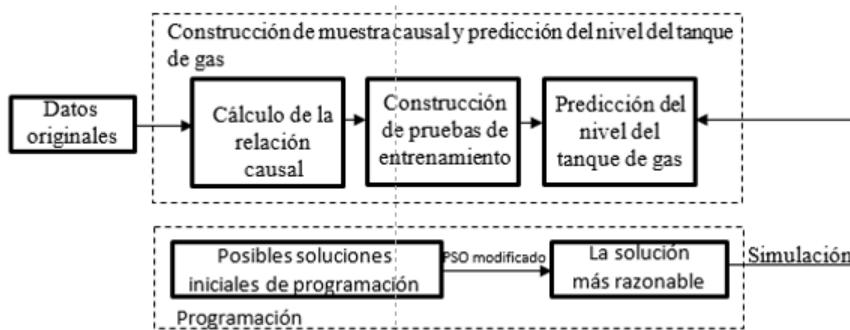


Figura 2. Modelo causal para el sistema CO.

Por otro lado, también se han encontrado estudios enfocados a la eficiencia energética, pero desde perspectivas de la utilización de recursos secundarios, los cuales son adquiridos o generados al interior de las compañías siderúrgicas, como los que se presentan a continuación:

- He y Wang en [20] sostienen que una de las alternativas que tienen las industrias siderúrgicas para mejorar la eficiencia de los recursos energéticos, es la utilización de energías secundarias, tales como el gas de carbón, calor residual y gases derivados de los procesos, de tal manera que estos recursos sean empleados como combustible en los procesos de producción del acero. Asimismo, también proponen una serie de sistemas industriales transversales (sistemas de caldera, energía distribuida / calor y energía combinados, etc.) que podrían apoyar la buena gestión de energéticos.
- Para Gao *et al.* en [5] la forma de lograr mejoras significativas en la gestión energética es con la aplicación de diversas tecnologías que incluyen el uso de inyección de carbón en los hornos, recuperación de calor residual sinterizado y utilización de deshumidificadores industriales de alto rendimiento, entre otras medidas.
- Por otro lado, en [7] Sun *et al.* ponen de manifiesto también una revisión exhaustiva de diversas tecnologías que pudieran apoyar en la optimización de la gestión energética. Dentro de las tecnologías mencionadas se encuentran aquellas que tienen que ver con el almacenamiento de energía, inyección de carbón pulverizado, utilización de calor residual y flujo de gas de los procesos industriales. Además, también proponen la introducción de modelos de pronóstico de flujos de materiales y energía para mejorar la eficiencia energética.
- Dentro de las tecnologías mencionadas por Hasan, Hoq y Thollander en [21] para mejorar la eficiencia energética se encuentran aquellas praxis que tienen que ver con la utilización de gases de escape y recuperación de calor de los procesos siderúrgicos, así como el empleo de inyección de carbón como uno de los combustibles en los procesos de fabricación.

3. Materiales y métodos

En base a los datos obtenidos del proceso en sitio, se seleccionó las técnicas de regresión para la generación de los modelos de predicción, siendo la Regresión Lineal Múltiple (RLM), la que mejor se adapta a los objetivos del presente estudio, debido a que este método estadístico permite examinar la relación y predicción de una variable respuesta cuantitativa en función de dos o más variables predictoras. Asimismo, permite identificar y clasificar de manera matemática qué variables explicativas son las que mayor impacto tienen sobre el fenómeno de estudio, cómo interactúan las variables entre sí y qué variables se pueden ignorar, entre otros aspectos. Y dada la naturaleza del fenómeno de estudio donde se tiene una variable dependiente y diversas variables explicativas asociadas a su comportamiento, se optó por tal técnica. Otro aspecto, el cual fue considerado para la selección de la técnica de predicción de RLM, es el número de observaciones obtenidas del fenómeno de estudio, el cual se ajustaba de forma apropiada para ser utilizado por este método. La metodología desarrollada para la presente investigación tiene un enfoque cuantitativo de tipo no experimental- transeccional, la cual tiene como objetivo la generación de un modelo matemático basado en la técnica predictiva de regresión lineal múltiple para estimar las unidades de consumo del volumen de gas natural en la industria acerera.

3.1. Descripción del modelo de regresión lineal múltiple

La técnica de regresión múltiple hace referencia a la relación de una variable dependiente con dos o más variables independientes, donde las variables independientes pueden combinar variables cuantitativas y cualitativas. Para denotar el número de variables independientes se suele usar p , tal como se muestra a continuación en la estructura del modelo de regresión lineal múltiple [22], [23]:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_p x_p \quad (1)$$

Donde \hat{y} = Es la variable dependiente, que representa el valor estimado a partir de una muestra. Y $b_0, b_1, b_2, \dots, b_p$ son las estimaciones de los parámetros $\beta_0, \beta_1, \beta_2, \dots, \beta_p$.

El análisis de regresión lineal múltiple, a diferencia del simple, se aproxima más a situaciones de análisis real puesto que los fenómenos hechos por definición, son complejos y, en consecuencia, deben ser explicados en la medida de lo posible por la serie de variables que directa e indirectamente participan en su concresción [24].

La regresión lineal múltiple posee seis supuestos claves que hay que considerar para realizar y desarrollar un análisis preciso y no sesgado durante la aplicación de la técnica [25], [26], [27]:

- Normalidad
- Ausencia de errores de medición
- Relación lineal
- La media de los residuales debe de ser igual a 0
- Multicolinealidad
- Homocedasticidad

3.1.1. Proceso de estimación en la regresión múltiple

Para realizar el proceso de regresión múltiple, se inicia con el cálculo de los parámetros desconocidos $b_0, b_1, b_2, \dots, b_p$, los cuales son las estimaciones de $\beta_0, \beta_1, \beta_2, \dots, \beta_p$. Una vez obtenidos los parámetros se integran al modelo de regresión lineal múltiple tal como se muestra en la Figura 3 [28].

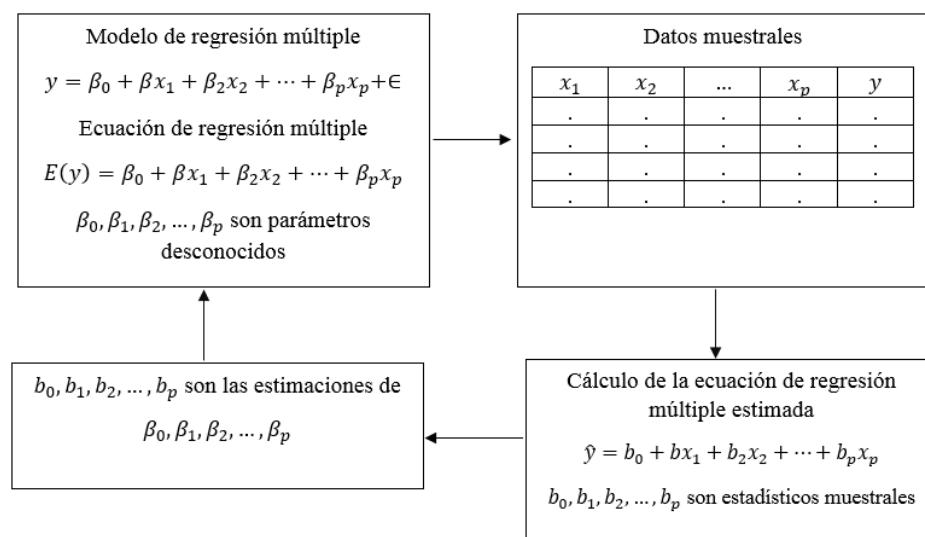


Figura 3. Procesos de estimación de parámetros en RLM.

3.1.2. Medidas de exactitud para la evaluación del modelo de regresión

Para evaluar el rendimiento de los modelos de regresión, se tienen unas series de medidas de exactitud entre las cuales destacan las siguientes [29]:

- Raíz del error cuadrático medio (RMSE): Es una medida absoluta de ajuste que es utilizada como criterio de predicción dentro de los modelos matemáticos, la cual representa el promedio del cuadrado de la

distancia entre el valor pronosticado y el valor real que se quiere estimar, tal como se muestra a continuación [30]:

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

Donde cada y_i corresponde a una observación confiable (\hat{y}_i) y n es el total de observaciones de un conjunto de valores.

- Error porcentual medio absoluto (MAPE): Métrica que tiene como objetivo medir el grado de certeza con que se realiza un valor estimado, midiéndose el tamaño del error, en términos porcentuales definiéndose de la siguiente manera [31]:

$$MAPE = \frac{100}{n} \times \sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right|, \quad (3)$$

- Desviación absoluta media (MAD): Medida estadística de error que representa el promedio de la distancia entre cada par de puntos reales y ajustados en un conjunto de datos y está dada por [32]:

$$MAD = \frac{\sum_{i=1}^n \left| \frac{x_i - \hat{x}_i}{x_i} \right|}{n}, \quad (4)$$

Donde: valor real en el periodo x_i ; \hat{x}_i es el valor de predicción en el periodo i ; y n es el número total de períodos temporales (número de observaciones).

- Coeficiente de determinación (r^2): Medida estadística para estimar el porcentaje de explicación de la variable dependiente a partir de una o más variables independientes, y su expresión es la siguiente [33]:

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

- Coeficiente de correlación (r): Es el encargado de medir la intensidad de la relación entre un conjunto de variables independientes y una variable dependiente, de tal forma que se puedan identificar las variables más influyentes sobre la variable dependiente. Los valores en los que puede oscilar el coeficiente de correlación son entre 1 y -1, donde 1 significa que existe una relación fuerte, y -1 una asociación fuerte pero negativa [34]. El coeficiente r se calcula de la siguiente manera:

$$r = \frac{S_{xy}}{S_x S_y}, \quad (6)$$

Donde S_{xy} es la covarianza muestral y $S_x S_y$ corresponde a la medida de variabilidad de la variable dependiente Y sin considerar el efecto de la variable independiente X .

3.2. Aplicación de la técnica predictiva regresión lineal múltiple

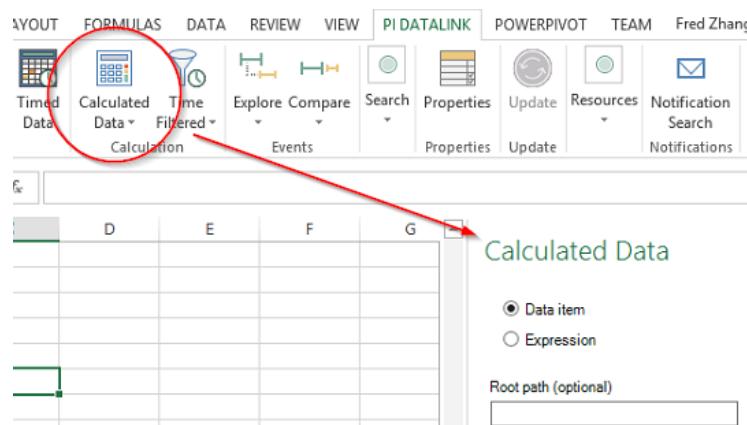
Se identificó una compañía dedicada a la actividad siderúrgica en México, con el fin de tener acceso y poder realizar un análisis exhaustivo al interior de la empresa, sobre los diversos mecanismos existentes e implementados para la administración y gestión de energéticos de gas natural dentro de la compañía. De acuerdo con el análisis del proceso en sitio, se tomó la decisión de considerar por departamento el volumen de consumo de gas natural, debido a la distinta forma de operar de cada uno de ellos, tomando como característica, que los departamentos seleccionados representan el grueso del consumo de gas natural de la empresa, obteniéndose cuatro departamentos dedicados a la actividad de producción y un departamento dedicado a la actividad de generación de energía. En la Tabla 1, se muestra la agrupación de los departamentos por similitud de actividad productiva y operación.

Tabla 1. Agrupación de departamentos por operación. Fuente: Elaboración propia.

Actividad	Operación	Departamento
Producción	1	Departamento A
Producción	1	Departamento B
Producción	2	Departamento D
Producción	2	Departamento E
Generación de Energía	3	Departamento C

3.2.1. Recopilación de datos

El proceso de recopilación de datos consistió en una serie de entrevistas no estructuradas, con los encargados de Información y Estadística Energética de la empresa acerera y las observaciones realizadas sobre el proceso en sitio. Se tuvo acceso al sistema informático encargado de recopilar y almacenar los datos generados por los diferentes departamentos de la compañía. Dentro de la información registrada por el sistema (Figura 4), se encuentran datos de producción, consumo de gas natural, consumo de gas coque e inyección de carbón, etc.

**Figura 4.** Sistema informático PI System.

Una vez descargadas las bases de datos históricas en formato Excel, se unificó en un solo archivo todo el conjunto de datos, para posteriormente clasificarlos según la actividad de cada departamento. Dentro de la actividad de producción se agruparon cuatro departamentos, mientras que en la actividad de generación de energía eléctrica se agrupó un departamento (Tabla 2). El período de descarga de datos comprendió del año 2014 al 2019, de cada uno de los departamentos.

Tabla 2. Clasificación de información según la actividad por departamento. Fuente: Elaboración propia.

Actividad	Departamento	Año	Mes	Día	Volumen	Volumen	Inyección
					de consumo de gas natural solicitado (m ³)	de consumo de gas natural real (m ³)	
Producción	Departamento D	2017	10	1	25380.00	33523.73	4556.98
Producción	Departamento D	2017	10	2	25380.00	33914.76	4464.21
Producción	Departamento D	2017	10	3	25380.00	34295.72	4594.89
Producción	Departamento D	2017	10	4	25380.00	34523.70	4379.66
Producción	Departamento D	2017	10	5	25380.00	29895.30	3707.81
Producción	Departamento D	2017	10	6	25380.00	29359.13	3658.07
Producción	Departamento D	2017	10	7	25380.00	34374.00	4198.31
							4455.09

Producción	Departamento D	2017	10	8	25380.00	29205.08	3594.24	3735.21
Producción	Departamento D	2017	10	9	25380.00	29762.55	4477.43	4985.54
Producción	Departamento D	2017	10	10	25380.00	28876.43	4541.26	5307.40
Producción	Departamento D	2017	10	11	25380.00	29080.73	4623.28	5830.69
Producción	Departamento D	2017	10	12	25380.00	29138.25	4618.86	5892.87
Producción	Departamento D	2017	10	13	25380.00	29143.58	4674.50	5998.08
Producción	Departamento D	2017	10	14	25380.00	20346.15	4557.05	5854.66
Producción	Departamento D	2017	10	15	25380.00	26601.68	4611.82	5858.34
Producción	Departamento D	2017	10	16	25380.00	25607.70	3531.51	4547.00
Producción	Departamento D	2017	10	17	25380.00	2810.03	1236.40	1407.69
Producción	Departamento D	2017	10	18	25380.00	25811.63	2961.84	3440.69
Producción	Departamento D	2017	10	19	25380.00	28278.00	4657.63	5702.15
Producción	Departamento D	2017	10	20	25380.00	28252.28	4556.50	5384.88
Producción	Departamento D	2017	10	21	25380.00	37661.25	4378.27	3596.37
Producción	Departamento D	2017	10	22	25380.00	37606.65	3930.44	0123.58
Producción	Departamento D	2017	10	23	00000.00	02883.38	0870.94	0000.00

3.2.2. Transformación y limpieza de datos

Una vez clasificada la información de los distintos departamentos, se procedió a identificar valores inválidos que pudieran sesgar la presente investigación; se inició con la eliminación de aquellos registros que tuvieran valores vacíos o negativos, que hacen referencia a una mala lectura por parte de los sensores, los cuales son los encargados de recopilar los datos de consumo de gas natural de los diversos procesos de la empresa.

Después se analizó el conjunto de datos para revisar la posible existencia de valores atípicos mediante la prueba de Dixon; esto consiste en determinar si el valor más grande o el valor más bajo de una muestra, o los dos valores más grandes, o los dos más pequeños pueden considerarse valores atípicos. Una vez aplicada la técnica de Dixon, se identificó la presencia de valores atípicos en cada conjunto de datos, tal como se puede observar según el *p-values* calculado en cada grupo de datos (Tabla 3). Los valores atípicos fueron eliminados del conjunto de datos a analizar.

Tabla 3. Prueba de Dixon para valores atípicos. Fuente: Elaboración propia.

Departamento	Número de observaciones	p-values>alfa
Departamento A	1747	0.653>0.05
Departamento B	1479	0.885>0.05
Departamento C	1759	0.082>0.05
Departamento D	1290	0.255>0.05
Departamento E	475	0.109>0.05

3.2.3. Identificación de variables

Con base en la información obtenida por el sistema informático y las entrevistas realizadas con cada uno de los departamentos de interés para modelar y caracterizar el fenómeno de estudio de este trabajo, se logró identificar las siguientes variables explicativas, las cuales fueron definidas con respecto a la actividad dedicada y a la forma de operación de cada departamento; la variable respuesta corresponde al volumen de consumo de gas natural. En la Tabla 4, se muestran las variables consideradas para cada modelo.

Tabla 4. Designación de las variables (de respuesta y explicativas) de los modelos. Fuente: Elaboración propia.

Actividad	Departamento	Variable respuesta	Variables predictoras
Producción	Departamento A	Volumen de consumo de gas natural (m ³)	Producción (t), eficiencia, volumen de gas coque (m ³) y horas de trabajo promedio de las máquinas.

Producción	Departamento B	<i>Volumen de consumo de gas natural (m³)</i>	<i>Producción (t), eficiencia, volumen de gas coque (m³) y horas de trabajo promedio de las máquinas.</i>
Producción	Departamento D	<i>Volumen de consumo de gas natural (m³)</i>	<i>Producción, masa de inyección de carbón (t), eficiencia y temperatura ambiente (°C).</i>
Producción	Departamento E	<i>Volumen de consumo de gas natural (m³)</i>	<i>Producción (t), masa de inyección de carbón (t), eficiencia y temperatura ambiente (°C).</i>
Generación de Energía	Departamento C	<i>Volumen de consumo de gas natural (m³)</i>	<i>Producción (MW), eficiencia y temperatura ambiente (°C)</i>

Nota: Las unidades de medidas utilizadas para las diferentes variables identificadas son: t = toneladas, m³ = metros cúbicos, °C = grados centígrados y MW = megavatio.

La selección de variables fue realizada conforme a lo siguiente: dadas la similitud de operación del departamento A y B se seleccionaron las siguientes variables:

1. La variable de *producción (t)* fue seleccionada para ambos departamentos (A y B) debido a la importancia del impacto que tiene en el consumo de gas natural por mes, asimismo, esta variable de acuerdo con el personal encargado de la programación energética resulta fundamental para que la empresa se pueda mantener competitiva frente a sus rivales, siempre y cuando puedan lograr producciones optimas con el mejor pronóstico de consumo de gas natural sin caer en excesos que proporcionen pérdidas económicas.
2. Variable de *eficiencia*, este indicador está relacionado directamente con el uso de volumen de consumo de gas natural de cada una de las unidades de producción inherente a los procesos siderúrgicos. Convirtiéndose así en un aspecto relevante para disminuir o incrementar el volumen de consumo de gas natural dentro de la compañía.
3. *Volumen de gas coque (m³)*, esta variable ha sido identificada por la importancia del impacto que tiene en el consumo de gas natural en los procesos siderúrgicos, debido a que mediante el uso de esta alternativa de combustible se puede controlar o disminuir de forma positiva el volumen de consumo de gas natural.
4. El funcionamiento de los equipos (*horas de trabajo promedio de las máquinas*) es uno de los factores decisivos en la productividad de la empresa que impacta de forma directa en el consumo de gas natural, debido a que cuando surgen algunos problemas (incontrolables por el personal operativo). Los encargados de la nominación del gas natural tienden a realizar peticiones de abastecimiento a su máxima capacidad para cada una de las máquinas en operación, provocando con ello el suministro excesivo de gas natural sin llegar a utilizarse en ocasiones.

Al igual que en los departamentos anteriores, se definieron las variables de *producción (t)* y *eficiencia*, sin embargo, para los departamentos D y E se agregaron las siguientes variables de análisis:

1. *Masa de inyección de carbón (t)*, esta variable fue seleccionada debido a que la inyección de combustible de este tipo en los departamentos D y E en la mayoría de los casos es utilizada para reducir el consumo de gas natural por tonelada producida.
2. La *temperatura ambiente* en específico para estos departamentos y de acuerdo con los encargados de la programación energética, el clima resultaba ser un factor importante en el consumo de gas natural. La importancia de esta variable radica en el hecho de que los usuarios de los departamentos (D y E) manifiestan que la maquinaria utilizada en el proceso de operación en los meses de invierno suele tardar más para obtener el calentamiento operativo del horno, provocando así un mayor consumo de volumen de gas natural. En cambio, en los meses donde la temperatura ambiente es alta, tiende a tardar menos el proceso de calentamiento del horno utilizados por los departamentos.

Finalmente se realizó la selección de variables del único departamento dedicado a la generación de energía, el cual, al igual que en los departamentos anteriores posee las mismas características de las variables de *producción (t)* y *eficiencia*. Asimismo, que en los departamentos de operación de tipo 2, para este departamento fue

considerada la variable de temperatura ambiente debido al incremento del uso de la maquinaria de operación del departamento E.

3.2.4. Construcción de los modelos de RLM

Para la construcción del modelado de los datos mediante la técnica de predicción de RLM, se utilizó Excel a través del complemento XLSTAT, la cual tiene incorporadas una serie de técnicas de modelado y funciones estadísticas para el análisis y comportamiento de datos. Se siguió el proceso que se muestra en la Figura 5 con tres fases; la primera corresponde a la selección de variables, después se calcularon las medidas de exactitud para determinar la validez y confiabilidad de cada uno de los modelos y finalmente se realizó la evaluación de estos modelos mediante dos grupos de entrenamiento; el primer grupo de datos correspondió a los registros históricos del proceso en sitio de dos meses, mientras que el segundo grupo de datos estuvo compuesto por los datos estimados (promedios) de las variables definidas para cada modelo.

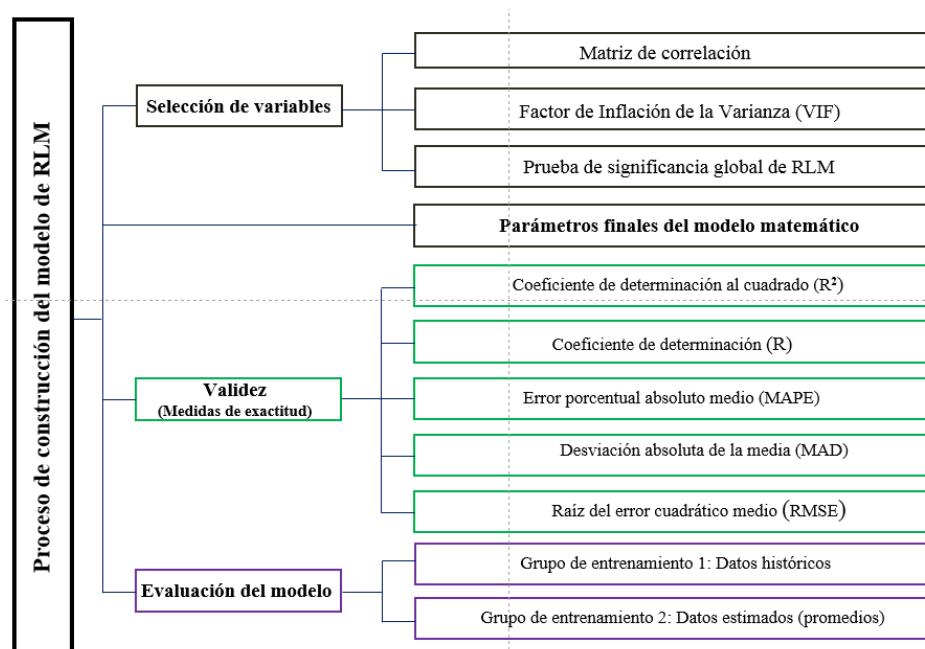


Figura 5. Construcción de los modelos de regresión lineal múltiple.

4. Resultados

Para validar y determinar si cada modelo propuesto para los diferentes departamentos (A, B, C, D y E) mejora el pronóstico de *volumen de consumo de gas natural*, se tomaron como referencia las medidas de exactitud de cada modelo que se muestran en la Tabla 5 y las medidas de error de pronóstico arrojadas por el grupo de datos de entrenamiento en la Tabla 6, donde se pudo observar que los indicadores de RMSE, MAD y MAPE de los modelos basados en regresión lineal múltiple, obtienen mejores resultados de ajustes con respecto al método tradicional utilizado por la compañía.

Tabla 5. Resultados obtenidos de las medidas de exactitud de cada uno de los modelos. Fuente: Elaboración propia.

Modelo	Ecuación	Medidas de exactitud		
		R ²	MAPE (%)	RMSE
Modelo 1 (Departamento A)	$Volumen\ de\ consumo\ de\ gas\ natural = 111953.01 - Volumen\ de\ gas\ coque \times 0.428 + Horas\ de\ trabajo\ promedio\ de\ las\ máquinas \times 11941.17$	0.737	5.222	22731.85

	<i>Volumen de consumo de gas natural =</i>				
	61958.11 + Producción × 52.53				
Modelo 2	+ Volumen de gas coque × 0.075				
(Departamento	- Eficiencia × 2502424.96	0.817	10.249	9451.62	
B)	-Horas de trabajo promedio de las máquinas				
	× 112.71				
	<i>Volumen de consumo de gas natural</i>				
	= 131107.881				
Modelo 3	+ Producción × 73.074				
(Departamento	- Masa de inyección de carbón	0.944	7.346	26864.53	
D)	× 7.772				
	- Eficiencia × 10131807.66				
	+ Temperatura ambiente				
	× 1412.169				
	<i>Volumen de consumo de gas natural</i>				
	= 289297.433				
Modelo 4	+ Producción × 98.971				
(Departamento	- Masa de inyección de carbón × 5.385 -	0.977	2.892	13797.77	
E)	Eficiencia × 27707698.14				
	+Temperatura ambiente × 557.167				
	<i>Volumen de consumo de gas natural</i>				
Modelo 5	= 215873.685				
(Departamento	+ Producción × 0.246	0.997	0.141	661.111	
C)	-Eficiencia × 532949.574 +				
	Temperatura ambiente × 7.290				

Nota: **R²**= Coeficiente de determinación al cuadrado, **MAPE** = Error porcentual absoluto medio y **RMSE**= Raíz del error cuadrático medio.

Tabla 6. Medidas de error de pronóstico de los diferentes modelos propuestos. Fuente: Elaboración propia.

	Modelo 1 RLM/MT	Modelo 2 RLM/MT	Modelo 3 RLM/MT	Modelo 4 RLM/MT	Modelo 5 RLM/MT
RMSE	35305.06 / 134958.70	10278.01 / 34581.12	48384.42 / 111661.65	27490.14 / 104146.06	187.19 / 23928.90
MAD	28292.73 / 101706.09	9178.78 / 30989.32	40649.32 / 90766.52	21767.43 / 79500.53	156.90 / 22892.10
MAPE (%)	9.44 / 37.52	15.68 / 49.33	17.02 / 44.80	8.90/ 25.94	0.09 / 12.67

Nota: Método de regresión lineal múltiple (RLM), método tradicional (MT), raíz del error medio al cuadrado (RMSE), desviación absoluta de la media (MAD), error porcentual absoluto medio (MAPE).

En el análisis de los modelos propuestos de regresión lineal múltiple no se observó presencia de multicolinealidad entre las variables explicativas de los modelos 2, 3, 4 y 5, debido a que los valores obtenidos de la prueba del Factor de Inflación de la Varianza (VIF) no fueron superiores a 10, por lo tanto, se consideraron todas las variables regresoras definidas para cada modelo [35]. Sin embargo, en el modelo 1 se presentó multicolinealidad en dos de sus variables, por lo cual fueron descartadas del modelo. En la Tabla 7 se muestran los resultados del VIF de cada modelo.

Tabla 7. Prueba del Factor de Inflación de la Varianza para los modelos de regresión lineal múltiple.

Modelo	Variables regresoras					
	Producción	Eficiencia	Volumen de gas coque	Horas de trabajo promedio de las máquinas	Masa de inyección de carbón	Temperatura ambiente
Modelo 1 (Departamento A)	70.785	1.047	1.174	71.794	NA	NA
Modelo 2 (Departamento B)	1.495	5.590	6.681	1.056	NA	NA
Modelo 3 (Departamento D)	VIF	2.455	1.299	NA	NA	2.513
Modelo 4 (Departamento E)		1.794	1.370	NA	NA	1.919
Modelo 5 (Departamento C)		1.455	1.430	NA	NA	1.102

Nota: NA= No aplica.

Fuente: Elaboración propia.

4.1. Pronósticos usando el modelo de regresión lineal múltiple

Con la aplicación de las técnicas de predicción de RLM para el departamento A, se puede observar en la Figura 6, que el método de RLM obtiene mejor resultado de pronóstico de *volumen de consumo de gas natural* con respecto al método tradicional, obteniendo un 9.44% de error porcentual absoluto medio (MAPE) con respecto al 37.52% obtenido por el método tradicional mostrado en la Tabla 6. Esto es respaldado por las medidas de exactitud calculadas para el Modelo 1, en el cual se determinó que las variables independientes explican la variabilidad del consumo de gas natural en un 74%, mientras que MAPE fue del 5.22%. Se confirma la importancia de la variable *horas de trabajo promedio de las máquinas* como uno de los factores más influyentes en el *volumen de consumo de gas natural* según el análisis de la suma de cuadrados tipo III.

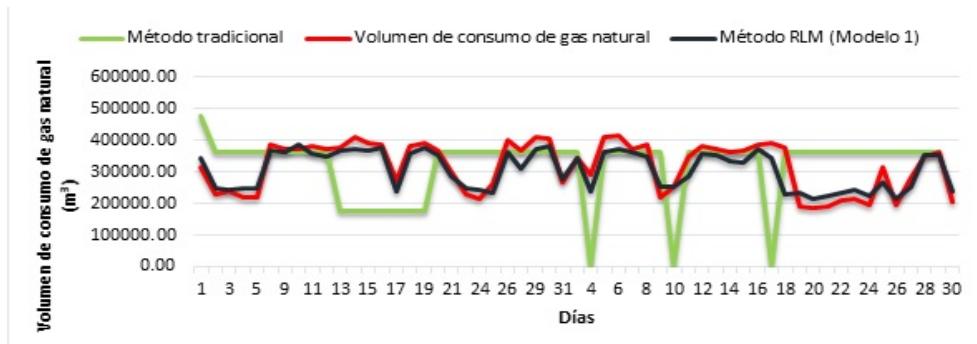


Figura 6. Comparación del método de RLM (Modelo 1) con respecto al método tradicional para el pronóstico del volumen de consumo de gas natural (m^3).

En la Figura 7 se muestran los resultados obtenidos para el departamento B, donde se determina que el método de RGMLM (Modelo 2) sigue manteniendo una mejor predicción del consumo de gas natural con respecto al método tradicional; se obtuvo un mejor porcentaje de error (MAPE) de 15.68% con respecto al método tradicional que obtuvo un 49.33%. Pese a que la naturaleza del departamento es compleja debido a diversos factores no controlables (días de paros no programables, defecto en alguna de sus máquinas de trabajo y accidentes imprevistos en el área operativa) que suelen influir de manera directa sobre el consumo de gas natural, las medidas de exactitud del Modelo 2 reflejan que las variables independientes para este modelo explican en 82% la variabilidad de la variable dependiente de *volumen de consumo de gas natural* de forma confiable, aportando así, pronósticos más

assertivos que el método tradicional utilizado actualmente. Para este modelo, el grupo de datos que obtuvo mejor comportamiento de pronóstico fue el grupo uno.

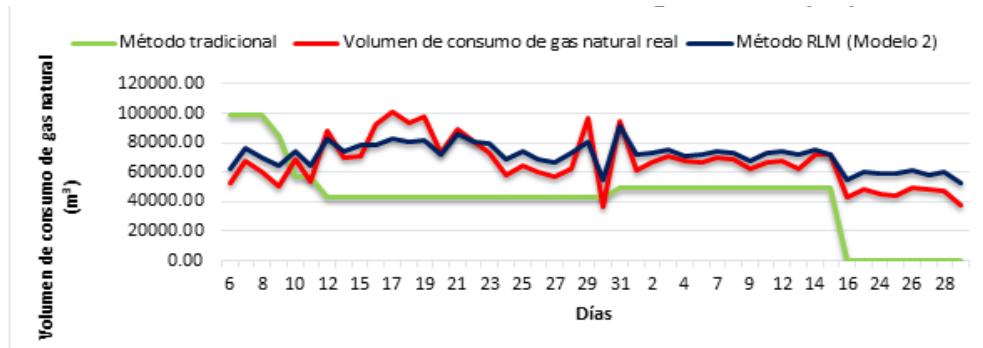


Figura 7. Comparación del método de RLM (Modelo 2) con respecto al método tradicional para el pronóstico del volumen de consumo de gas natural (m^3).

Los resultados obtenidos del método de RLM para el departamento D se muestran en la Figura 8, donde se observa que en este método fue incorporada la variable exógena de *temperatura ambiente*, la cual permitió mejorar de manera positiva las predicciones sobre el volumen de consumo de gas natural para el departamento. El error porcentual absoluto medio del Modelo 3 (basado en RLM) obtuvo un valor de 17.02%, mientras que el método tradicional un 44.80%, proporcionando así un modelo moderadamente factible para hacer pronóstico sobre la variable dependiente (*volumen de consumo de gas natural*). Al igual que en los modelos anteriores, las medidas de exactitud del modelo 3 respaldan tales resultados, donde se pudo determinar que las cuatro variables independientes (*producción, masa de inyección de carbón, eficiencia y temperatura ambiente*) explican en un 94 % la variabilidad de la variable dependiente.

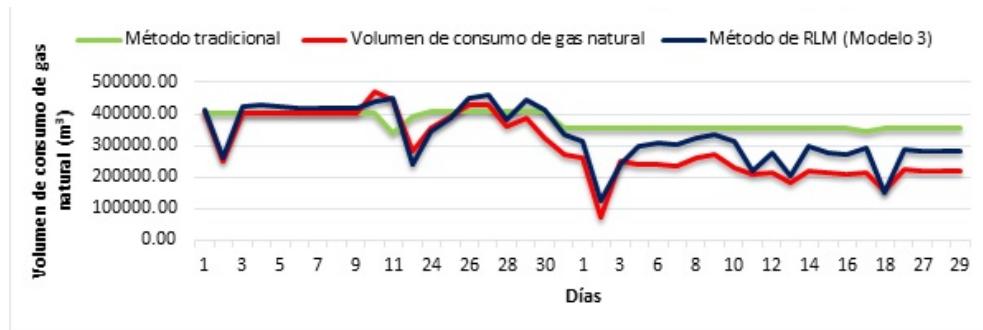


Figura 8. Comparación del método de RLM (Modelo 3) con respecto al método tradicional para el pronóstico del volumen de consumo de gas natural (m^3).

Para el Modelo 4 del departamento E se asume que, este método proporciona una mejor predicción de la variable dependiente con respecto al método tradicional (Figura 9), donde el valor del error porcentual absoluto medio correspondió al 8.90% y 25.94% respectivamente. También se identificó que las variables independientes en su conjunto explican en un 98% la variabilidad del *volumen de consumo de gas natural*. Asimismo, se determinó que la variable *eficiencia* se relaciona e influye más con el consumo de gas natural en el departamento E.

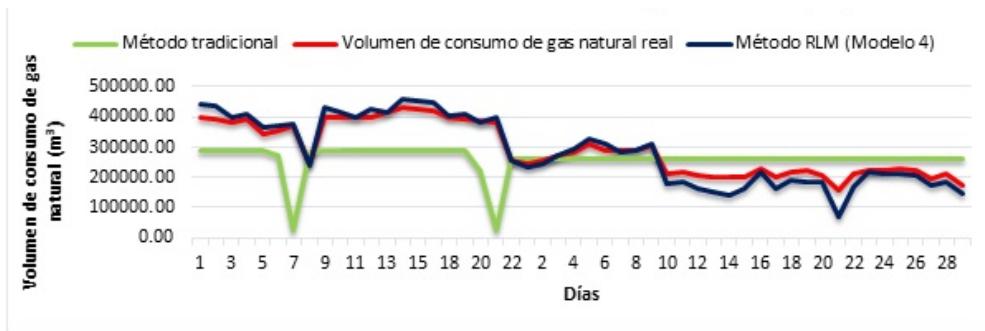


Figura 9. Comparación del método de RLM (Modelo 4) con respecto al método tradicional para el pronóstico del volumen de consumo de gas natural (m^3).

A continuación, se muestran los resultados obtenidos del método de RLM (Modelo 5) del departamento C, dedicado a la actividad productiva de energía eléctrica dentro de la compañía. Para este modelo, se tuvieron como variables independientes: *producción, eficiencia y temperatura ambiente*; mientras que la variable dependiente fue el *volumen de consumo de gas natural*. Este modelo presentó los mejores resultados respecto a los modelos anteriores, obteniendo como resultados de MAPE de 0.09% con respecto al método tradicional con un 12.67%; mientras que el R^2 para el Modelo 5 fue de 0.997, respaldando así la factibilidad para hacer proyecciones más assertivas que el método tradicional sobre consumo de gas natural (Figura 10). Entre las variables explicativas del Modelo 5 que más influyen sobre la variable dependiente se encuentra la variable de *producción*, siendo esta la de mayor correlación con respecto a las demás variables.

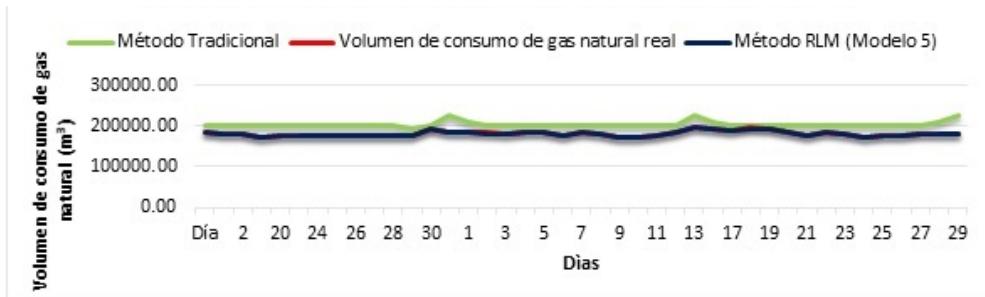


Figura 10. Comparación del método de RLM (Modelo 5) con respecto al método tradicional para el pronóstico del volumen de consumo de gas natural (m^3).

5. Discusión de resultados

A partir de los resultados y del análisis de las medidas de exactitud de cada uno de los modelos propuestos, se determinó que, mediante el uso de un modelo matemático de gestión de recursos de gas natural, se pueden hacer estimaciones más precisas y confiables que permitan mejorar la eficiencia de los recursos energéticos en la industria siderúrgica en México.

Los resultados de las medidas de exactitud de los cinco modelos generados para la variable dependiente de *volumen de consumo de gas natural* con sus respectivas variables explicativas obtuvieron correlaciones fuertes (Modelo 1: 0.85, Modelo 2: 0.90, Modelo 3: 0.97, Modelo 4: 0.98, Modelo 5: 0.99). Por otra parte, los modelos que mostraron mejor comportamiento de pronóstico sobre la variable dependiente fueron los Modelos 1, 4 y 5 con un MAPE de 9.44%, 8.9% y 0.09% respectivamente. Con las variables identificadas para mejorar las estimaciones de consumo de gas natural en los modelos 3, 4 y 5, se considera que la variable exógena es la temperatura ambiente, debido a que mediante tal variable los procesos siderúrgicos pueden regular y estimar mejor el pronóstico de consumo de gas natural dentro de la compañía.

Con base en los resultados del análisis de la suma de cuadrados de tipo III, en los Modelos 2, 3 y 5 la variable que más influyó en el *volumen de consumo de gas natural* fue la variable independiente *producción*; mientras que para el Modelo 1 fue la variable de *horas de trabajo promedio de las máquinas* y para el Modelo 4 fue la variable *eficiencia*.

Los resultados anteriormente explicados fueron obtenidos del análisis del proceso en sitio y conforme a los datos proporcionados por los encargados de la parte operativa de la unidad de análisis, teniendo como limitación el factor de confidencialidad en los datos crudos del proceso de consumo de gas natural en cada uno de los procesos siderúrgicos.

6. Conclusiones

Se ha identificado que actualmente la industria siderúrgica carece de herramientas de gestión de recursos energéticos, provocando graves problemas en la optimización de consumo de gas natural en sus diversos procesos industriales. Sin embargo, este problema puede ser abordado mediante el uso de modelos predictivos basados en regresión lineal múltiple, los cuales, según las medidas de exactitud calculadas, proporcionan mejores pronósticos sobre el volumen de consumo de gas natural que los métodos convencionales actuales.

Los resultados del error porcentual absoluto medio y el error cuadrático medio de los modelos de RLM (MAPE: 10.23%, RMSE: 20492.32, DAM: 20009.03) vs. método tradicional (MAPE: 34.05%, RMSE: 93055.92, DAM: 65170.91), reflejan que los modelos propuestos mejoran la gestión de recursos de gas natural, debido a que proporcionan estimaciones más confiables y precisas del volumen de consumo de gas natural en los procesos siderúrgicos, de tal forma que se puedan realizar planificaciones adecuadas de la programación energética que conlleve a un incremento en su productividad, ahorros monetarios y eficiencia en el manejo de sus recursos energéticos de gas natural.

Por otro lado, se encontró, que a medida que los departamentos utilizaban recursos secundarios como combustible en sus procesos, tales como: volumen de gas coque o masa de inyección de carbón, estos disminuían considerablemente el volumen de consumo de gas natural en los procesos siderúrgicos.

7. Referencias

- [1] Viteri Rade, L. Y., Franco Villon, M. N. (2022). El desarrollo organizacional a través del talento humano. *E-IDEA Journal of Business Sciences*, 4 (17), 30-44. <https://doi.org/10.53734/eidea.vol4.id233>
- [2] Hu, Z., He, D., Feng, K., Liu, P., Jia, Y. (2019). Optimal Design Model of the Energy Systems in Iron and Steel Enterprises. *Applied Sciences*, 9 (22), 2-9. <https://doi.org/10.3390/app9224778>
- [3] Wang, Y., Wen, Z., Yao, J., Doh Dinga, C. (2020). Multi-objective optimization of synergic energy conservation and CO₂ emission reduction in China's iron and steel industry under uncertainty. *Renewable and Sustainable Energy Reviews*, 134, 1-13. <https://doi.org/10.1016/j.rser.2020.110128>
- [4] Yuan, Y., Na, H., Chen, C., Qiu, Z., Sun, J., Zhang, L., Du, T., Yang, Y. (2024). Status, challenges, and prospects of energy efficiency improvement methods in steel production: A multi-perspective review. *Energy*, 304. <https://doi.org/10.1016/j.energy.2024.132047>
- [5] Gao, C., Gao, W., Song, K., Na, H., Tian, F., Zhang, S. (2019). Comprehensive evaluation on energy-water saving effects in iron and steel industry. *Science of the Total Environment*, 670, 346-360. <https://doi.org/10.1016/j.scitotenv.2019.03.101>
- [6] Wang, X., Zhang, T., Luo, S., Abedin, M. Z. (2023). Pathways to improve energy efficiency under carbon emission constraints in iron and steel industry: Using EBM, NCA and QCA approaches. *Journal of Environmental Management*, 348. <https://doi.org/10.1016/j.jenvman.2023.119206>
- [7] Sun, W., Wang, Q., Zhou, Y., Wu, J. (2020). Material and energy flows of the iron and steel industry: Status quo, challenges and perspectives. *Applied Energy*, 268, 1- 15. <https://doi.org/10.1016/j.apenergy.2020.114946>
- [8] Nenchev, B., Panwisawas, C., Yang, X., Fu, J., Dong, Z., Tao, Q., Gebelin, J., Dunsmore, A., Dong, H., Li, M., Tao, B., Li, F., Ru, J., Wang, F (2022). Metallurgical Data Science for Steel Industry: A Case Study on Basic Oxygen Furnace. *Steel research international*, 93, 1-13. <https://doi.org/10.1002/srin.202100813>
- [9] Karthick, K., Dharmapakash, R., Sathya, S. (2024). Predictive Modeling of Energy Consumption in the Steel Industry Using CatBoost Regression: A Data-Driven Approach for Sustainable Energy Management. *International Journal of Robotics and Control Systems*, 4 (1), 33-49. <https://doi.org/10.31763/ijrcs.v4i1.1234>

- [10]Rojas, J. C., Hasanbeigi, A., Sheinbaum, C., Price, L. (2017). Energy efficiency in the Mexican iron and steel industry from an international perspective. *Journal of Cleaner Production*, 158, 335-348. <https://doi.org/10.1016/j.jclepro.2017.04.092>
- [11]Wolniak, R., Saniuk, S., Grabowska, S., Gajdzik, B. (2020). Identification of Energy Efficiency Trends in the Context of the Development of Industry 4.0 Using the Polish Steel Sector as an Example. *Energies*, 13 (11), 2-16. <http://dx.doi.org/10.3390/en13112867>
- [12]Zhou, P., Xu, Z., Zhu, X., Zhao, J., Song, C., Shao, Z. (2024). Granulation-based long-term interval prediction considering spatial-temporal correlations for gas demand prediction in the steel industry. *Expert Systems with Applications*, 248. <https://doi.org/10.1016/j.eswa.2024.123382>
- [13]Wang, R. Q., Jiang, L., Wang, Y. D., Roskilly, A. P. (2020). Energy saving technologies and mass-thermal network optimization for decarbonized iron and steel industry: A review. *Journal of Cleaner Production*, 274, 1-28. <https://doi.org/10.1016/j.jclepro.2020.122997>
- [14]Norbert, R., Kim, J., Griffay, G. (2020). A system dynamics framework for the assessment of resource and energy efficiency in iron and steel plants. *Journal of Cleaner Production*, 276, 1-10. <https://doi.org/10.1016/j.jclepro.2020.123663>
- [15]Ahmad, I., Arif, M. S., Cheema, I. I., Thollander, P., Khan, M. A. (2020). Drivers and Barriers for Efficient Energy Management Practices in Energy-Intensive Industries: A Case-Study of Iron and Steel Sector. *Sustainability*, 12 (18), 1-16. <https://doi.org/10.3390/su12187703>
- [16]Karimi-Zare, A., Shakouri G, H., Kazemi, A., Kim, E.-S. (2024). Aggregate production planning and energy supply management in steel industry with an onsite energy generation system: A multi-objective robust optimization model. *International Journal of Production Economics*, 269. <https://doi.org/10.1016/j.ijpe.2024.109149>
- [17]Tang, L., Meng, Y. (2021). Data analytics and optimization for smart industry. *Frontiers of Engineering Management*, 8, 157–171. <https://doi.org/10.1007/s42524-020-0126-0>
- [18]Díaz, D., Ocampo, O. (2022). *Gas natural para la transición energética y competitividad de México*. Instituto Mexicano para la Competitividad A. C. <https://imco.org.mx/wp-content/uploads/2022/08/Gas-Natural-Competitivo-en-Mexico.pdf>
- [19]Jin, F., Lv, Z., Li, M., Mou, L., Zhao, J., Wang, W. (2018). A Causal Model-Based Scheduling Approach for Coke Oven Gas System in Steel Industry. *IFAC-PapersOnLine*, 51 (21), 7-12. <https://doi.org/10.1016/j.ifacol.2018.09.384>
- [20]He, K., Wang, L. (2017). A review of energy use and energy-efficient technologies for the iron and steel industry. *Renewable and Sustainable Energy Reviews*, 70, 1022-1039. <https://doi.org/10.1016/j.rser.2016.12.007>
- [21]Hasan, M., Hoq, T., Thollander, P. (2018). Energy management practices in Bangladesh's iron and steel industries. *Energy Strategy Reviews*, 22, 230-236. <https://doi.org/10.1016/j.esr.2018.09.002>
- [22]Maaouane, M., Zouggar, S., Krajacic, G., Zahboune, H. (2021). Modelling industry energy demand using multiple linear regression analysis based on consumed quantity of goods. *Energy*, 225. <https://doi.org/10.1016/j.energy.2021.120270>
- [23]Soto-Bravo, F., González-Lutz, M. I. (2019). Análisis de métodos estadísticos para evaluar el desempeño de modelos de simulación en cultivos hortícolas. *Agronomía Mesoamericana*, 30 (2), 517-534. <https://doi.org/10.15517/am.v30i2.33839>
- [24]Izhar, I., Radiman, R., Wahyun, S. F. (2023). Investigating factors affecting employees' work productivity. *Journal of Enterprise and Development (JED)*, 5 (Special-Issue-2), 1-18. <https://journal.uinmataram.ac.id/index.php/jed/article/view/7982>
- [25]Vilà Baños, R., Torrado Fonseca, M., Reguant Álvarez, M. (2019). Análisis de regresión lineal múltiple con SPSS: un ejemplo práctico. *Revista d'Innovació i Recerca en Educació*, 12 (2), 1-10. <http://doi.org/10.1344/reire2019.12.222704>
- [26]Alita, D., Putra, A. D., Darwis, D. (2021). Analysis of Classic assumption test and multiple linear regression coefficient test for employee structural office recommendation. *Indonesian Journal of Computing and Cybernetics Systems*, 15 (3), 295-306. <https://doi.org/10.22146/ijccs.65586>
- [27]Pisica, D., Dammers, R., Boersma, E., Volovici, V. (2022). Tenets of Good Practice in Regression Analysis. A Brief Tutorial. *World Neurosurgery*, 161, 230-239. <https://doi.org/10.1016/j.wneu.2022.02.112>
- [28]Levin, R., Rubin, D. (2004). *Estadística para administración y economía* (7ma. Ed.). Pearson Educación.

- [29] Mejía Vásquez, E. J., Gonzales Chávez, S. (2019). Predicción del consumo de energía eléctrica residencial de la Región Cajamarca mediante modelos Holt -Winters. *Ingeniería Energética*, XL (3), 181-191.
<https://www.redalyc.org/journal/3291/329160723002/html/>
- [30] Robeson, S. M., Willmott, C. J. (2023). Decomposition of the mean absolute error (MAE) into systematic and unsystematic components. *Plos one*, 18 (2), 1-18. <https://doi.org/10.1371/journal.pone.0279774>
- [31] Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development*, 15 (14), 5481–5487. <https://doi.org/10.5194/gmd-15-5481-2022>
- [32] Voloh, B., Watson, M. R., Konig, S., Womelsdorf, T. (2020). MAD saccade: statistically robust saccade threshold estimation via the median absolute deviation. *Journal of Eye Movement Research*, 12 (8), 1-12. <https://doi.org/10.16910/jemr.12.8.3>
- [33] Rodríguez Sánchez, A., Salmerón Gómez, R., García, C. (2019). The coefficient of determination in the ridge regression. *Communications in Statistics - Simulation and Computation*, 51 (1), 201-219. <https://doi.org/10.1080/03610918.2019.1649421>
- [34] Schober, P., Boer, C., Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia & Analgesia*, 126 (5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>
- [35] Vega Vilca, J. C., Guzman, J. (2011). Regresión PLS y PCA como solución al problema de multicolinealidad en regresión múltiple. *Revista de Matemática Teoría y Aplicaciones*, 18 (1), 9–20. <https://www.redalyc.org/articulo.oa?id=45326927002>