



Un enfoque basado en Machine Learning para analizar la influencia de la inversión del canon minero en el índice de desarrollo humano en el departamento de Cajamarca

A Machine Learning-based approach to analyze the influence of mining royalty investment on the human development index in the department of Cajamarca

Rosel Burga Cabrera¹, José Camilo Micha Ortiza¹, Néstor Elías Muñoz Abanto¹, Paul Cueva Araujo¹

¹Facultad de Ingeniería, Universidad Nacional de Cajamarca, Cajamarca, Perú

Resumen

El impacto de la inversión en el canon minero en el Índice de Desarrollo Humano (IDH) en el departamento de Cajamarca ha sido examinado minuciosamente mediante un enfoque fundamentado en el Aprendizaje Automático. Los hallazgos sugieren que, pese a las considerables inversiones efectuadas entre 2008 y 2019, no se ha detectado una correlación relevante entre el ingreso per cápita proveniente del canon minero y el IDH, con un coeficiente de correlación de 0.0062. Al analizar los elementos del Índice de Desarrollo Humano, se descubrió que tanto la expectativa de vida como el nivel de educación no muestran vínculos relevantes con las inversiones del canon. Específicamente, para la educación de jóvenes de 18 a 25 años, los coeficientes de correlación fueron de 0.043 y -0.023, respectivamente, lo que indica que la inversión no ha generado un efecto beneficioso en la educación. Igualmente, el estudio comparativo entre los distritos muestra que la diversidad en la inversión del canon minero no resulta en incrementos justos en los indicadores de desarrollo humano, ingresos o educación. Se determinó que la regresión lineal es el modelo más apropiado para esta investigación, evidenciando el error medio cuadrático más bajo ($MSE = 25294.503$) y un RMSE de 159.042, lo cual señala su eficacia predictiva. Estos descubrimientos subrayan la imperiosa necesidad de una administración estratégica del canon minero que privilegie el bienestar social y fomente un progreso justo en Cajamarca. Esto implica que la mera distribución de recursos no asegura avances en el crecimiento humano sin una correcta organización y un enfoque holístico en su implementación.

Palabras clave

Canon minero, índice de Desarrollo Humano, Machine learning



Abstract

The impact of mining royalty investment on the Human Development Index (HDI) in the department of Cajamarca has been thoroughly examined using a Machine Learning-based approach. The findings suggest that despite the considerable investments made between 2008 and 2019, no relevant correlation has been detected between per capita income from mining royalties and the HDI, with a correlation coefficient of 0.0062. When analyzing the elements of the Human Development Index, it was found that both life expectancy and education level do not show relevant links with royalty investments. Specifically, for the education of young people aged 18 to 25, the correlation coefficients were 0.043 and -0.023, respectively, indicating that the investment has not generated a beneficial effect on education. Likewise, the comparative study between districts shows that diversity in mining royalty investment does not result in fare increases in human development, income or education indicators. Linear regression was determined to be the most appropriate model for this research, showing the lowest mean square error ($MSE = 25294.503$) and an RMSE of 159.042, which indicates its predictive effectiveness. These findings underline the imperative need for strategic management of the mining royalty that privileges social well-being and encourages fair progress in Cajamarca. This implies that the mere distribution of resources does not ensure progress in human growth without proper organization and a holistic approach in its implementation.

Keywords

Mining canon, Human Development Index, Machine learning

Introducción

El Perú se caracteriza porque su economía está basada en la explotación de recursos naturales. Dicha circunstancia ha conllevado a que las actividades extractivas ganen un notable protagonismo dentro de la estructura productiva nacional, sobre todo de las actividades mineras. En el departamento de Cajamarca, esta realidad resulta muy evidente, pues la actividad minera es fundamental para la región y ha sido un motor clave para el desarrollo económico. Cajamarca es conocida por sus ricas reservas minerales, incluyendo oro, plata y cobre, lo que ha atraído tanto inversiones nacionales como extranjeras.

Las inversiones en el sector minero han permitido la implementación del canon minero, un mecanismo de distribución de recursos generado por esta actividad. Este canon consiste en



un porcentaje de las regalías que las empresas mineras deben pagar al Estado, el cual es posteriormente redistribuido a los gobiernos locales y regionales. Esto ha permitido que Cajamarca reciba importantes ingresos que pueden ser utilizados para financiar proyectos de infraestructura, educación, salud y desarrollo social.

En este contexto, es necesario establecer la manera en la que el canon minero influye en el desarrollo humano dentro de la región de Cajamarca y evaluar los resultados obtenidos. Para ello, se empleará un modelo de Machine Learning que permita predecir la relación entre el canon minero y el desarrollo humano, lo cual puede ofrecer información valiosa que facilite la toma de decisiones y la planificación de políticas públicas.

El objetivo de la presente investigación es utilizar técnicas de Machine Learning para analizar datos socioeconómicos y demográficos recolectados a nivel distrital en todo el departamento de Cajamarca, para construir un modelo predictivo que permita conocer la relación entre la inversión del canon minero y el índice del desarrollo humano. Empleando esta información se busca generar políticas públicas efectivas y eficientes para mejorar las condiciones de vida de la población cajamarquina y reducir las brechas de desigualdad existentes. Además, Para el desarrollo del modelo predictivo, se utilizarán técnicas de Machine Learning. Estos métodos permitirán identificar patrones y relaciones entre las variables estudiadas, así como predecir el impacto del canon minero en el índice del desarrollo humano.

Los resultados obtenidos del modelo se presentarán de manera clara y accesible, con el fin de que tanto los tomadores de decisiones como la sociedad civil puedan comprender las implicaciones de las inversiones del canon minero en el desarrollo humano. Se buscará fomentar un diálogo entre todos los actores involucrados para asegurar que las políticas públicas formuladas a partir de esta investigación sean inclusivas y estén alineadas con las necesidades y aspiraciones de la población cajamarquina.

Material y métodos

La presente investigación se desarrolló mediante la aplicación de técnicas de minería de datos y aprendizaje automático (Machine Learning), con el objetivo de analizar la relación entre la implementación de programas sociales y el control de la anemia infantil en el departamento de Cajamarca. El enfoque metodológico integró procesos de recopilación, preprocesamiento, modelado y evaluación de datos, siguiendo estándares propios de la analítica predictiva.



Recopilación y preprocesamiento de datos

En una primera etapa, se realizó una recopilación sistemática de información relacionada con los programas sociales y su posible impacto en la anemia infantil en Cajamarca. Esta incluyó datos demográficos, indicadores socioeconómicos, información sobre la cobertura e implementación de programas sociales, así como estadísticas de prevalencia de anemia infantil. Las fuentes de información provinieron de registros oficiales, bases de datos institucionales y literatura especializada.

Posteriormente, se llevó a cabo un proceso riguroso de limpieza de datos, orientado a garantizar la calidad y consistencia del conjunto de datos. Este proceso incluyó la identificación y tratamiento de valores faltantes, la eliminación de registros duplicados, la depuración de datos inconsistentes y la exclusión de variables irrelevantes para el análisis. Asimismo, se optimizó la estructura del dataset mediante la eliminación de columnas redundantes.

En la fase de transformación, los datos fueron adaptados a los requerimientos del modelo de Machine Learning. Esto implicó la normalización de variables numéricas, la codificación de variables categóricas y la generación de nuevas características (feature engineering) que permitieron enriquecer el conjunto de datos y mejorar la capacidad predictiva del modelo.

Selección y entrenamiento del modelo

Para el modelado, se seleccionaron diversos algoritmos de Machine Learning adecuados para problemas de predicción y análisis de relaciones complejas entre variables. Entre los algoritmos considerados se incluyeron árboles de decisión, regresión lineal, k-vecinos más cercanos (k-NN), máquinas de soporte vectorial (SVM) y redes neuronales artificiales, cada uno con ventajas específicas en términos de interpretabilidad, capacidad predictiva y manejo de datos no lineales.

El conjunto de datos fue dividido en dos subconjuntos: uno destinado al entrenamiento del modelo y otro para su evaluación. Esta partición permitió estimar el rendimiento del modelo en datos no observados durante el entrenamiento, reduciendo el riesgo de sobreajuste (overfitting).

Durante la fase de entrenamiento, los modelos ajustaron sus parámetros internos utilizando el conjunto de entrenamiento, con el objetivo de minimizar el error de predicción. Cabe señalar que, a diferencia de versiones preliminares del diseño, el modelo fue orientado



específicamente a predecir la prevalencia de anemia infantil en función de variables asociadas a programas sociales, corrigiendo inconsistencias conceptuales previas relacionadas con variables no pertinentes como el índice de desarrollo humano o el canon minero.

Evaluación y validación del modelo

Una vez entrenados, los modelos fueron evaluados utilizando el conjunto de prueba. Se emplearon métricas cuantitativas como el error cuadrático medio (MSE), así como otros indicadores de desempeño según la naturaleza del modelo (por ejemplo, precisión, recall o coeficiente de determinación R^2 en modelos de regresión), con el fin de medir la capacidad predictiva y la generalización de los algoritmos.

En cuanto a los materiales utilizados, el análisis se apoyó en herramientas de software especializadas en ciencia de datos. Se emplearon plataformas como Orange Data Mining para el análisis exploratorio y la visualización, así como el lenguaje de programación Python para el desarrollo de modelos más avanzados. En este entorno, se utilizaron bibliotecas como scikit-learn y TensorFlow, fundamentales para la implementación, entrenamiento y evaluación de algoritmos de Machine Learning.

El procesamiento de datos se llevó a cabo en equipos de cómputo con capacidad suficiente para manejar operaciones de análisis y modelado, garantizando la eficiencia en los cálculos. Adicionalmente, se revisó literatura científica y documentación técnica relevante sobre anemia infantil y programas sociales en contextos similares, lo cual permitió contextualizar los resultados y fortalecer la interpretación de los hallazgos.

En conjunto, este enfoque metodológico integral asegura la robustez del análisis y la validez de los resultados obtenidos, proporcionando una base sólida para la comprensión de la relación entre programas sociales y el control de la anemia infantil en Cajamarca.

Resultados y discusión

El estudio realizado examina el impacto de la inversión del canon minero en el Índice de Desarrollo Humano (IDH) en la región Cajamarca, Perú. El canon minero, como mecanismo de redistribución de recursos, tiene como objetivo beneficiar a las regiones productoras mediante la financiación de proyectos de desarrollo social e infraestructura. Teniendo en cuenta que Cajamarca es una de las principales regiones mineras del país, recibe una parte significativa de estos fondos, el equipo de investigación. El estudio tiene como objetivo brindar un análisis basado en datos que ayude a las autoridades a comprender mejor la relación entre los recursos



del canon minero y el progreso social y económico en la región desde la perspectiva de la gestión pública. Este enfoque posibilitará la identificación de patrones ocultos en los datos históricos y simular el impacto potencial de nuevas inversiones, contribuyendo a una asignación más eficiente de los fondos y a un desarrollo más equitativo y sostenible en Cajamarca.

Comprensión de los datos

En el presente estudio los datos utilizados fueron obtenidos del Ministerio de Economía y Finanzas (MEF), específicamente de registros oficiales que detallan la inversión del canon minero, así como del Programa de las Naciones Unidas para el Desarrollo (PNUD) que proporciona información sobre el Índice de Desarrollo Humano (IDH) y factores socioeconómicos relevantes. El análisis se fundamenta en técnicas de machine learning avanzadas para explorar la relación entre estas inversiones y el IDH, lo que permite identificar patrones y tendencias que no son evidentes a simple vista. Se incluyen datos de inversión asignados anualmente a Cajamarca, detallados por áreas como infraestructura, educación y salud, lo cual resulta fundamental para comprender su impacto potencial en el desarrollo regional. Asimismo, se incorporarán datos sobre el Índice de Desarrollo Humano, el cual mide factores como la salud, la educación y el nivel de vida. Este enfoque integral permitió soslayar el impacto específico del canon minero en el desarrollo humano y ofrecer recomendaciones para optimizar la gestión de estos recursos en Cajamarca.

Exploración de los datos

En la figura 1 nos muestra la evolución de los promedios de ingresos per cápita desde los años 2008 hasta 2019, también el promedio de ingresos correspondientes al canon minero del mismo período y la variación del índice de desarrollo humano.

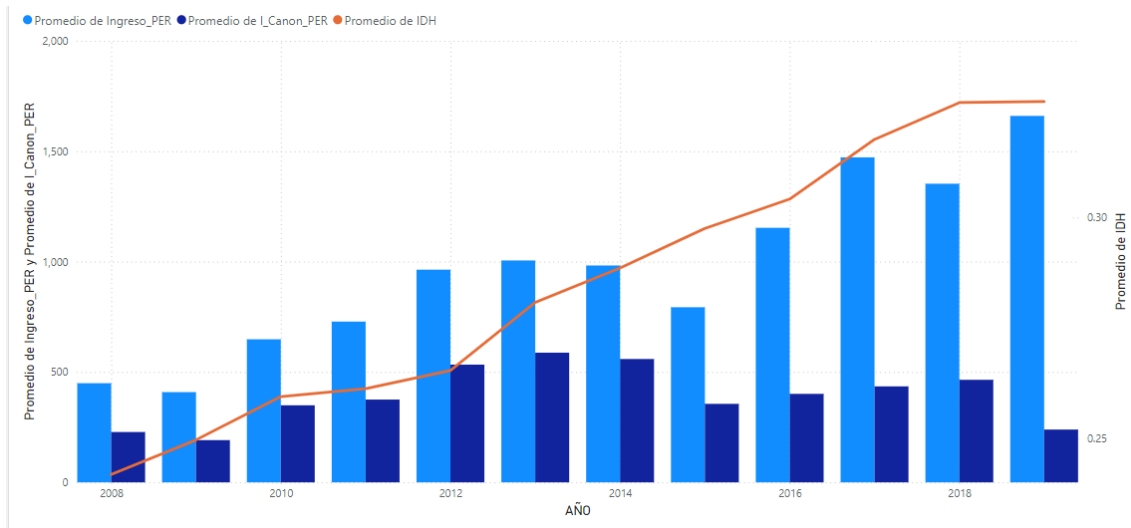


Figura 1. Variación de los ingresos per cápita municipales y el índice de desarrollo humano del departamento de Cajamarca para los períodos 2008 - 2019.

En la figura 1 se ilustra la variación del presupuesto de ingresos de los gobiernos municipales del departamento de Cajamarca durante el periodo comprendido entre 2008 y 2019. Este análisis revela un incremento significativo en el presupuesto total, aunque es importante destacar que la asignación de recursos provenientes del canon minero mostró un patrón diferente. Tras un crecimiento sostenido hasta 2013, se observó una tendencia a la baja en estos ingresos que continuó hasta 2019. A pesar de esta fluctuación en los ingresos por canon minero, el Índice de Desarrollo Humano (IDH) de la región experimentó un aumento constante durante el mismo periodo. Estos resultados sugieren la ausencia de una correlación directa entre los ingresos generados por el canon minero y el avance del desarrollo humano en el departamento de Cajamarca, lo que plantea interrogantes sobre la efectividad de la inversión de estos recursos en la mejora de la calidad de vida de la población local.

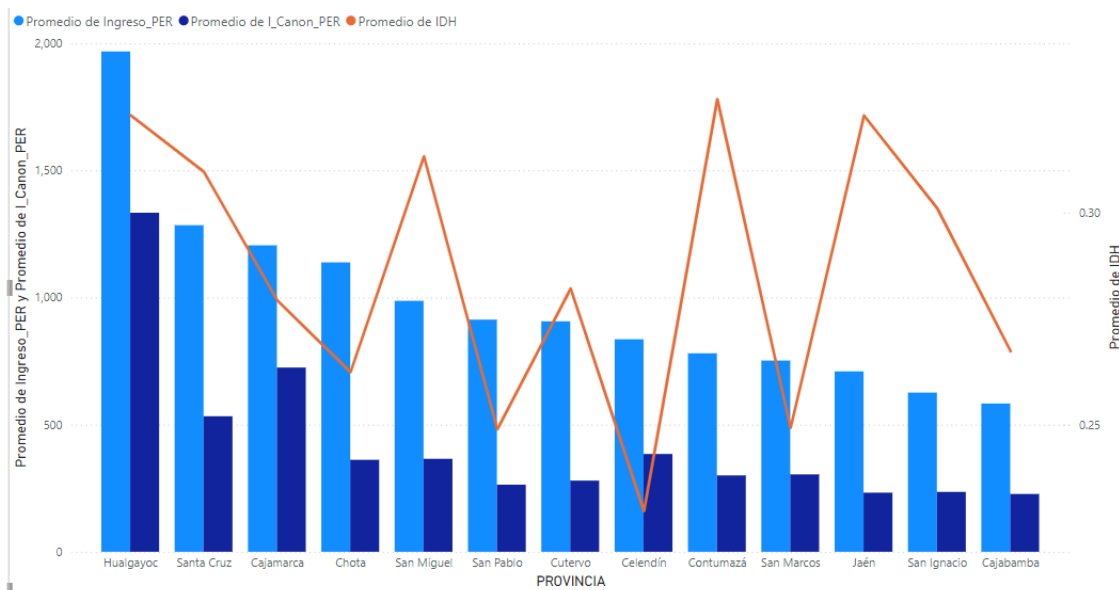


Figura 2. Variación de los ingresos per cápita municipales y el índice de desarrollo humano del departamento de Cajamarca para los períodos 2008 - 2019.

En la figura 2 se presentan los ingresos generados por concepto de canon minero en el presupuesto de cada una de las provincias del departamento de Cajamarca. Este análisis revela una notable discrepancia entre los niveles de ingreso provenientes del canon minero y los promedios del Índice de Desarrollo Humano (IDH) en estas localidades. En particular, se observa que las provincias de Cajamarca, Jaén y San Miguel presentan índices de desarrollo humano elevados, lo que sugiere un bienestar social relativamente alto. Sin embargo, a pesar de este desempeño en el IDH, estas mismas provincias registran ingresos bajos en el rubro del canon minero dentro de sus presupuestos municipales, en este contexto se plantean cuestiones sobre la distribución y uso de los ingresos mineros, así como su impacto en el desarrollo socioeconómico de las comunidades locales, lo que podría indicar la necesidad de una revisión crítica de las políticas asociadas a la explotación de recursos minerales en la región.

Preparación de los datos

Es esencial considerar la preparación adecuada de los datos con el fin de asegurar la excelencia y precisión de los modelos de Machine Learning en el análisis de la influencia de la inversión del canon minero en el Índice de Desarrollo Humano (IDH) en el departamento de Cajamarca. Este procedimiento consta de múltiples fases que garantizan la eficacia del análisis y la modelización de los datos, adaptándolos a las particularidades de la investigación. En primer lugar, se transforma la inversión del canon minero en una inversión per cápita,



dividiendo las cifras totales por distrito entre la población correspondiente, lo cual posibilita una comparación equitativa del impacto de las inversiones.

$$\text{Inversión per cápita}_{\text{distrito,año}} = \frac{\text{Inversión total en el distrito, año}}{\text{Población del distrito, año}}$$

En segundo lugar, se atribuye cualquier dato faltante del IDH mediante un modelo de regresión lineal, ya que este indicador tiende a variar con rapidez en el tiempo.

$$\text{IDH estimado}_t = \beta_0 + \beta_1 \cdot \text{Año}_t + \epsilon_t$$

Por último, se procede a la estandarizar y normalizar los datos con el fin de garantizar la convergencia de todas las variables en los modelos de Machine Learning. Asimismo, se lleva a cabo una detección y gestión de los outliers con el fin de prevenir la distorsión de los resultados. En última instancia, se agrupan todos los datos relevantes en un conjunto final que incluye la inversión per cápita, los valores imputados del IDH y variables socioeconómicas, seguido de una división en conjuntos de entrenamiento y prueba con el fin de garantizar la validez del modelo. Este enfoque integral permitirá optimizar el impacto de las inversiones del canon minero en el desarrollo humano en Cajamarca.

Modelado

La preparación de datos es un proceso clave que nos permite asegurar la calidad y precisión de nuestros modelos de aprendizaje automático para analizar el impacto de las inversiones mineras en el Índice de Desarrollo Humano (IDH) de Cajamarca. El proceso anterior implica varios pasos para asegurar que los datos estén preparados para el análisis y el modelado adaptados a las características específicas del estudio, por lo que primero hay que convertir la tasa de rendimiento de la inversión en inversión per cápita dividiendo el número total por región por la población correspondiente, lo que permite una comparación justa del impacto de la inversión. En segundo lugar, se utiliza un modelo de regresión lineal para imputar los datos faltantes del IDH, ya que este indicador tiende a cambiar lentamente con el tiempo. Luego, los datos se estandarizan y normalizan para garantizar que todas las variables sean comparables en el modelo de aprendizaje automático. Además, se pueden detectar y gestionar valores atípicos para evitar la distorsión de los resultados. Finalmente, todos los datos relevantes se combinan en un conjunto final que incluye la inversión per cápita, los valores estimados del IDH y las variables socioeconómicas y luego se dividen en conjuntos de



capacitación y prueba para garantizar la validez del modelo para determinar el impacto que tuvo la inversión del canon y su impacto en el índice de desarrollo humano.

La regresión lineal es un clásico método estadístico que modela entre variables dependientes e independientes, caracterizándose por su sencillez y su interpretación fácil, aunque limitada en su capacidad para modelar relaciones que no son lineales. Por su parte, los métodos de Machine Learning, árboles de decisión, SVM, kNN, Random Forest y Gradient Boosting entre otros, demuestran tener ventajas importantes a la hora de trabajar con datos complejos, ya que les permite modelar interacciones no lineales y son robustos al sobreajuste siendo muy útiles para trabajar con datos que comprenden muchas características. Estas técnicas avanzadas permitirán la aplicación de modelos para llevar a cabo un análisis más completo y exacto del impacto que tienen las inversiones del canon minero en el IDH, facilitando así la implementación de políticas informadas y eficaces.

La elección del modelo más apropiado para este análisis debe tener en cuenta no solo la exactitud en la predicción, sino también la habilidad para entender y su importancia en el proceso de toma de decisiones. El uso de la regresión lineal en conjunto con métodos de aprendizaje automático puede proporcionar un entendimiento completo de cómo la inversión en el canon minero afecta el Índice de Desarrollo Humano (IDH) en Cajamarca. El propósito final es formular sugerencias fundamentadas en pruebas que favorezcan el progreso sostenible de la región.

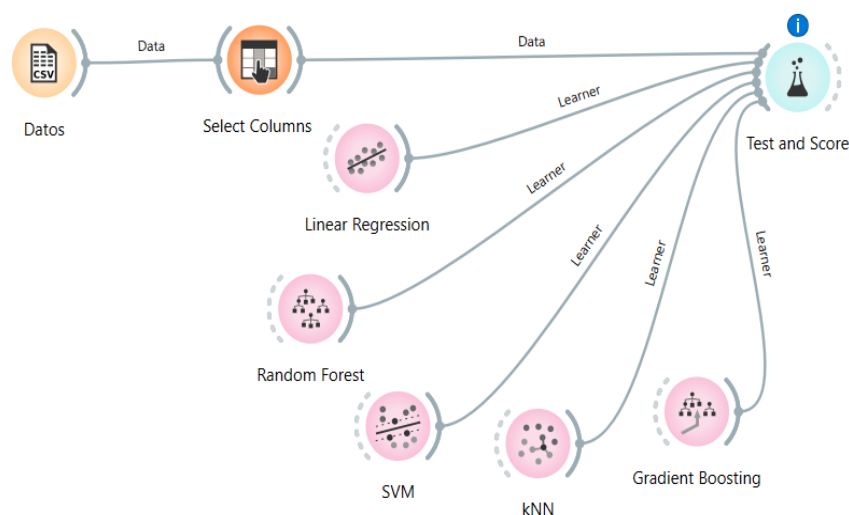


Figura 3. Flujo de trabajo para la valoración de modelos ideales en IDH y la inversión en la minería de CANON mediante Orange Data Mining.



Evaluación del modelo

Para medir el desempeño de los modelos, se utilizarán indicadores como el error medio cuadrático (MSE) y el coeficiente de determinación (R2), lo que facilitó la determinación de la capacidad predictiva de cada modelo.

Interpretación: Ya que el estudio no solo aspira a predecir, sino también a interpretar los hallazgos, la regresión lineal puede resultar beneficiosa debido a su claridad y sencillez para transmitir los resultados a los interesados, que incluyen a los responsables políticos y a la comunidad en su conjunto.

Tabla 1

Métricas para evaluar los modelos de machine learning

Modelo	MSE	RMSE	MAE	MAPE	R2
Linear Regression	25294.503	159.042	119.866	0.429	0.108
Random Forest	33162.387	182.105	134.810	0.468	-0.170
SVM	28001.784	167.337	134.598	0.519	0.012
Gradient Boosting	25702.821	160.321	119.177	0.422	0.093
kNN	28873.125	169.921	127.436	0.448	-0.019

Error cuadrático medio (MSE). Se utiliza frecuentemente en el aprendizaje automático para calcular la diferencia media cuadrática entre los valores esperados y reales de una variable.

$$MSE = \left(\frac{1}{n}\right) \sum_{j=1}^n (y_j - \hat{y}_j)^2$$

Para determinar el MSE, se requirió de datos que contiene valores reales conocidos y los correspondientes valores estimados. El MSE se determinó mediante la media de las variaciones al cuadrado entre los valores estimados y los reales. El valor medio de error cuadrático obtenido en la tabla 1 se corresponde con la regresión lineal (MSE = 25294.503) y (RMSE = 159.042), lo que sugiere que este sería el modelo que debemos escoger.



Coefficientes de correlación

Ppto Ingreso	1	0.95	0.1	0.34	0.048	0.25	0.33	0.34	0.15	0.22
Ingreso Canon	0.95	1	0.25	0.23	0.039	0.15	0.26	0.23	0.13	0.32
Pjte Canon	0.1	0.25	1	-0.22	-0.035	-0.25	-0.057	-0.22	-0.078	0.54
IDH	0.34	0.23	-0.22	1	0.074	0.76	0.8	0.95	0.13	0.0062
EspVida	0.048	0.039	-0.035	0.074	1	-0.049	-0.18	0.047	0.01	0.03
Ed_18_%	0.25	0.15	-0.25	0.76	-0.049	1	0.58	0.59	0.19	0.043
Ed_25_%	0.33	0.26	-0.057	0.8	-0.18	0.58	1	0.67	0.01	-0.023
PER	0.34	0.23	-0.22	0.95	0.047	0.59	0.67	1	0.13	-0.0011
Ingreso_PER	0.15	0.13	-0.078	0.13	0.01	0.19	0.01	0.13	1	0.6
I_Canon_PER	0.22	0.32	0.54	0.0062	0.03	0.043	-0.023	-0.0011	0.6	1

Figura 4. Coeficiente de correlación entre las distintas variables que están en investigación.

Entre 2008 y 2019, la inversión del canon minero en Cajamarca, calculada por el ingreso per cápita, no presenta una correlación relevante con el Índice de Desarrollo Humano (IDH). Este hallazgo se respalda en un coeficiente de correlación de 0.0062, lo que sugiere la inexistencia de correlación entre ambas variables. Al desglosar los componentes del IDH, se observa que la esperanza de vida y el ingreso per cápita derivado del canon presentan un coeficiente de correlación de 0.03, lo que refuerza la conclusión de que tampoco existe una relación significativa. En cuanto a la relación entre el ingreso per cápita de los habitantes y el ingreso per cápita de las municipalidades distritales, se reporta un coeficiente de -0.0011, indicando igualmente la falta de correlación entre estas variables, con referencia al factor educativo, se obtuvieron coeficientes de correlación de 0.043 y -0.023, que se alinean con la media de educación de los jóvenes de 18 y 25 años, respectivamente, lo que indica que no existe una correlación relevante entre la inversión del canon minero y el nivel de educación en estos colectivos. En resumen, estos estudios demuestran que, pese a la inversión en equipo



minero, no se han logrado avances notables en los indicadores de desarrollo humano, ingresos y educación en los distritos de Cajamarca durante el periodo analizado.

Se evaluó la eficacia de los modelos empleando indicadores como el error medio cuadrático (MSE) y el coeficiente de determinación (R^2) como resultados del análisis efectuado, lo que permitió establecer la capacidad predictiva de cada modelo. La interpretación de los resultados es igual de vital; por ende, la regresión lineal emerge como una alternativa beneficiosa gracias a su claridad y sencillez para transmitir los descubrimientos a los interesados, que incluyen a políticos y comunidades locales.

El efecto del canon minero en el Índice de Desarrollo Humano ha sido examinado en varias investigaciones, poniendo de manifiesto una relación complicada entre la inversión pública y el progreso social; en este ámbito, son múltiples estudios que proporcionan un panorama que facilita la comprensión del contexto de estos estudios. Velarde (2017), en su estudio de las regiones de Cajamarca y Ancash, utilizó una metodología econométrica basada en paneles de datos con efectos constantes. Sus descubrimientos mostraron que las transferencias del canon minero influyeron de manera considerable en el Índice de Desarrollo Humano de Ancash, en particular en el sector del transporte. No obstante, en Cajamarca no se registró un impacto significativo, lo que indica que la eficacia de las inversiones puede fluctuar considerablemente dependiendo de la región y la utilización particular de los recursos otorgados.

De forma similar, Rodríguez y Sotomayor (2020) utilizaron un método similar para examinar el caso de la provincia de Espinar, determinando que el canon minero tiene un impacto positivo en el Índice de Desarrollo Humano, resaltando un incremento significativo en áreas vitales como la educación y la salud. No obstante, también indicaron que el desembolso asignado al medio ambiente evidenció un impacto negativo, lo que resalta la importancia de un enfoque balanceado en la distribución de recursos Escarcena (2023) empleó un modelo econométrico de panel para la macro región sur, concluyendo que el canon minero ha aportado de manera notable a la mejora del IDH, lo que sugiere que una administración correcta de estos recursos es esencial para potenciar sus ventajas sociales.

Lozano (2024) llevó a cabo un estudio más extenso, analizando varias regiones de Perú y subrayando que, pese a que la industria minera es esencial para la economía del país, el efecto en el progreso humano no siempre se refleja en incrementos en los índices de desarrollo humano.



Estas investigaciones subrayan la relevancia de tener en cuenta no solo el monto de la inversión, sino también su adecuada repartición y uso en sectores vitales como la educación y la salud. En este contexto, los hallazgos del análisis efectuado sobre la inversión en el canon minero en Cajamarca durante el período 2008-2029, muestran una situación alarmante, las inversiones realizadas no se reflejan en un aumento del índice de desarrollo humano. Aunque estas inversiones generan expectativas, la correlación entre el ingreso per cápita proveniente del canon minero y el IDH se evidencia como insignificante, con un coeficiente de correlación de 0.0062 que señala la ausencia de correlación entre ambos índices. Al descomponer los elementos del Índice de Desarrollo Humano, se nota que la esperanza de vida y el ingreso per cápita exhiben un coeficiente de correlación de 0.03, corroborando la conclusión de la ausencia de una relación relevante.

Respecto al factor educativo, los coeficientes obtenidos para el promedio de educación de jóvenes de 18 a 25 años son de 0.043 y -0.023, respectivamente, lo que sugiere que no existe una correlación relevante entre la inversión del canon minero y el nivel de educación en estos grupos. De igual forma, al evaluar el ingreso per cápita de los residentes en relación con el ingreso per cápita de las municipalidades distritales, se registra un coeficiente de -0.0011, lo que también indica una falta de correlación entre estas variables.

Estos análisis conjuntos demuestran que, a pesar de la considerable inversión en el canon minero, no se han observado avances significativos en los indicadores de desarrollo humano, ingresos o educación en los distritos de Cajamarca durante el periodo analizado. Esto resalta la importancia de una administración estratégica del canon minero que privilegie el bienestar social y fomente un crecimiento justo en las zonas impactadas.

Por último, es importante señalar que, para valorar el desempeño de los modelos analíticos, se emplearán indicadores como el error medio cuadrático (MSE) y el coeficiente de determinación (R^2). Estas medidas facilitarán la determinación de la habilidad para predecir de cada modelo. Específicamente, la regresión lineal emerge como una alternativa beneficiosa gracias a su claridad y sencillez para transmitir los descubrimientos a los interesados, que incluyen a políticos y comunidades locales. La tabla 1 sintetiza las métricas de desempeño de los modelos estudiados: El estudio de estas métricas evidencia que el modelo de regresión lineal es el más apropiado para los objetivos de la investigación, presentando el MSE (25294.503) más bajo y el RMSE (159.042) más bajo, lo que señala su efectividad en la predicción de los resultados obtenidos.



Conclusiones

Hay una correlación insignificante entre la inversión y el crecimiento humano: Aunque el canon minero generó una considerable inversión en Cajamarca entre 2008 y 2019, se descubrió que la relación entre el ingreso per cápita producido por este y el Índice de Desarrollo Humano es casi inexistente (coeficiente de 0.0062). Esto implica que las inversiones no resultan inmediatamente en incrementos en los indicadores de desarrollo humano, lo que demanda una revisión de cómo se administran y reparten estos recursos para satisfacer las demandas de los habitantes.

El estudio de los elementos del Índice de Desarrollo Humano demostró que no hay una correlación relevante entre la inversión en el canon minero y el nivel de educación, además de la expectativa de vida en Cajamarca. Los coeficientes logrados para la educación de los jóvenes y el ingreso per cápita de los residentes señalan que, pese a las expectativas, las inversiones no han incrementado significativamente los grados de educación ni la calidad de vida de los habitantes.

La aplicación de algoritmos de aprendizaje automático, especialmente la regresión lineal, ha probado ser eficaz para estimar el Índice de Desarrollo Humano basándose en la inversión del canon minero y otras variables socioeconómicas. Este modelo, que cuenta con un error medio cuadrático (MSE) de 25294.503, se muestra como una herramienta apropiada para entender la conexión entre la inversión y el progreso humano.

Los descubrimientos subrayan la relevancia de una administración estratégica del canon minero que considere no solo el monto de la inversión, sino también su reparto justo y su uso en sectores vitales como la educación y la salud. Esto permitirá maximizar los beneficios sociales de dichas inversiones. Resulta imprescindible que los líderes políticos formulen políticas que alineen los objetivos de desarrollo humano con la distribución de recursos del canon minero.

Al examinar de forma comparativa los diferentes distritos de Cajamarca, se demuestra que la inversión del canon minero ha experimentado una variación significativa en su impacto en el progreso humano. Aunque algunos distritos podrían haber notado avances, otros han observado un efecto negativo o nulo en sus índices de IDH. Esto señala la importancia de un enfoque más individualizado en la distribución de recursos, considerando las circunstancias locales y promoviendo un crecimiento justo en toda la región.



Referencias

- Antoci, A., Russu, P., & Ticci, E. (2009). The impact of mining on local development: Evidence from Peru. *Resources Policy*, 123-132.
- Canavos, G. (2023). *Probabilidad y Estadística. Aplicaciones y Métodos*. México: McGraw-Hill.
- Chugh, A. (08 de Diciembre de 2020). *MAE, MSE, RMSE, coeficiente de determinación, R cuadrado ajustado: ¿qué métrica es mejor?* Obtenido de <https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e>
- Cruz, M., & Teixeira, A. (2010). Mining and local development: The case of Peru. *Journal of Development Studies*, 345-367.
- Escarcena, M. (2023). *Impacto del canon minero en el Índice de Desarrollo Humano de la macro región sur: 2003-2019*. Puno: Universidad Nacional del Altiplano.
- IBM. (17 de Agosto de 2021). *Guía de CRISP-DM de IBM SPSS Modeler*. Obtenido de <https://www.ibm.com/docs/es/spss-modeler/saas?topic=dm-crisp-help-overview>
- Kaminki, B., Jakubczyk, M., & Szufel, P. (2017). A framework for sensitivity analysis of decision trees. *Central European Journal of Operations Research*, 135–159.
- Loayza, N., & Rigolini, J. (2016). *The impact of mining on poverty and inequality in Peru*. World Bank Policy Research Working Paper.
- Lozano, D. (2024). *Efecto de los ingresos de canon en el bienestar de las regiones del Perú, 2004-2021*. Lima: Pontificia Universidad Católica del Perú.
- PNUD. (2010). *Human Development Report 2010: The real wealth of nations: Pathways to human development*. Programa de las Naciones Unidas para el Desarrollo.
- PNUD. (2015). *Human Development Report 2015: Work for human development*. Programa de las Naciones Unidas para el Desarrollo.
- PNUD. (2019). *Human Development Report 2019: Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century*. Programa de las Naciones Unidas para el Desarrollo.
- PNUD. (2020). *Human Development Report 2020: The next frontier: Human development and the anthropocene*. Programa de las Naciones Unidas para el Desarrollo.
- Rodriguez, L., & Sotomayor, J. (2020). *Influencia del canon minero en el índice de desarrollo humano de la provincia de Espinar, periodo 2015 – 2019*. Cusco: Universidad Nacional de San Antonio de Abad de Cusco.



- Ticci, E., & Escobal, J. (2015). Mining and its impact on employment and poverty in Peru. *International Journal of Social Economics*, 1-20.
- Velarde, L. (2017). *El impacto del canon minero en el índice de desarrollo humano a través de los gastos públicos en saneamiento, educación y transporte, 2010-2012: caso Ancash y Cajamarca*. Lima: Universidad de Lima.
- Webb, G., Boughton, J., & Wang, Z. (2005). Not So Naive Bayes: Aggregating One-Dependence Estimators. *Machine Learning*, 5-24.
- Zhan, J., & Vaswani, N. (2015). ACP robusto con conocimiento parcial del subespacio. *IEEE Transactions on Signal Processing*, 3332-3347.