



Universidad Tecnológica de Panamá
Facultad de Ingeniería de Sistemas Computacionales
Maestría en Analítica de Datos
Modelo Predictivos

Nombre de la investigación:

"Predicción de la productividad ciclo a ciclo de camiones de acarreo en minería usando modelos de regresión"

Nombre:

Liz Pérez

Grupo:

1AN-216

Facilitador:

Juan Marcos Castillos, PhD

23 de abril de 2025

INTRODUCCIÓN

En las industrias de movimiento de tierra, uno de los recursos más importantes es el tiempo: mientras más rápido se mueve el material, mayores ingresos se generan y se mejora la eficiencia del sistema. Factores operativos como la disponibilidad, utilización y productividad determinan las toneladas movidas al mes, lo cual incide directamente en los costos unitarios y en el cumplimiento de KPIs (indicadores clave de rendimiento). La incertidumbre en la cantidad de toneladas producidas o el incumplimiento de la meta puede generar pérdidas millonarias.

En este contexto, la minería moderna se enfrenta al reto de aprovechar el gran volumen de datos generados por la operación diaria para mejorar la toma de decisiones. Este proyecto propone un análisis predictivo para estimar la productividad promedio de camiones (en toneladas por hora) a partir de variables operacionales clave como la carga útil y la duración del ciclo. Se implementan y comparan distintos modelos de regresión, tanto lineales como basados en árboles de decisión, para evaluar su capacidad de predicción y su aplicabilidad práctica.

JUSTIFICACIÓN

La productividad de los camiones mineros depende de múltiples factores, como el tipo de material transportado, los tiempos de espera, la duración de carga, descarga y trayecto. Comprender estos factores y modelarlos permite identificar cuellos de botella, reducir tiempos improductivos y maximizar el uso de la maquinaria. Al utilizar herramientas de ciencia de datos y aprendizaje automático sobre datos reales, se puede transformar la toma de decisiones operativas en acciones basadas en evidencia cuantificable.

ANTECEDENTES

Uno de los principales retos en las operaciones de acarreo minero es anticipar la variabilidad en la productividad de los camiones, especialmente cuando pequeños retrasos en los ciclos de viaje pueden tener impactos muy significativos en la producción. Desde que inicié a trabajar en la industria, he estado expuesta a este tipo de desafío que se relacionan con la planificación operativa. En más de una ocasión he visto cómo decisiones importantes se toman basadas en promedios históricos o simplemente en la intuición, sin un respaldo analítico fuerte.

Cada ciclo de acarreo representa: tiempo, combustible, desgaste del equipo y toneladas por mover. Sin embargo, aunque las operaciones recopilan una gran cantidad de datos operacionales, rara vez o casi nunca se utilizan para predecir comportamientos futuros. Se pierde así una oportunidad clave para mejorar la eficiencia y anticiparse a problemas. Normalmente los parámetros utilizados para la planificación tienen que ver con valores ya predeterminados o de softwares minero que no toman en cuenta la realidad operacional.

He trabajado en proyectos de optimización, mejora continua y análisis de KPIs, y siempre he creído que el uso de modelos predictivos es un paso natural hacia una minería más inteligente. Este proyecto nace de esa convicción y del interés genuino de aplicar mis conocimientos técnicos en un problema que vivo de cerca, combinando mi experiencia en operaciones mineras con mi formación en ciencia de datos.

A nivel internacional, hay referentes claros sobre cómo el análisis predictivo puede transformar decisiones tácticas y estratégicas en minería. Sin embargo, en Panamá y en muchas operaciones de la región centroamericana, aún estamos en proceso de adoptar estas tecnologías de forma integrada. Por eso, esta propuesta busca ser un caso práctico y realista, que sirva como ejemplo de lo que se puede lograr con datos bien estructurados y modelos bien implementados.

DEFINICIÓN DEL PROBLEMA

En una operación minera, los camiones de acarreo cumplen un ciclo cuando depositan el material que cargan en su destino, ya sea una trituradora o un acopio. Un ciclo está constituido por: viaje vacío hacia la pala, tiempo de cola, tiempo de aculatamiento, tiempo de carga de material en la pala, viaje lleno hacia el destino, tiempo de cola, tiempo de aculatamiento, tiempo de descargar y se repite el proceso. La duración de cada uno de estos ciclos, sumada a la cantidad de material transportado, define cuán productivo es un camión por hora. Esta productividad está sujeta a múltiples variaciones operativas como cambios de material, condiciones climáticas, granulometría del material, distancias, entre otras.

Esto implica que, aunque un camión tenga un ciclo largo, si lleva una carga significativamente mayor, podría compensar ese tiempo. Y viceversa: un ciclo corto con baja carga podría resultar en baja productividad. Ante estas fluctuaciones, surge la necesidad de predecir con precisión la productividad promedio (ton/hora) en base a variables disponibles en la operación, especialmente para tomar decisiones en tiempo real.

El desafío radica en construir un modelo predictivo capaz de anticipar la productividad con base en los datos operativos más representativos, de forma que los supervisores y planificadores puedan anticipar desviaciones, redirigir recursos y optimizar la asignación de equipos antes de que se vea afectada la producción.

ANÁLISIS PREDICTIVO

a) Determinación de la base de datos

Para el desarrollo del proyecto se utilizó un conjunto de datos operacionales de un sistema que recopila los datos de los eventos de carguío de las palas eléctricas la cual llena unos camiones mineros, en una operación minera a cielo abierto. Cada registro incluye variables como hora del evento, ID de equipo, duración del ciclo, ubicación, material extraído, tiempo de espera, y otros indicadores clave que permiten caracterizar la eficiencia del proceso.

La base de datos fue seleccionada por contar con una granularidad adecuada y representar situaciones operativas reales en distintos momentos del turno, permitiendo analizar el comportamiento de la flota en distintas condiciones de operación y turnos. El objetivo principal es identificar patrones que permitan predecir la duración de los ciclos y, con ello, mejorar la planificación del despacho de camiones y optimizar el rendimiento de la flota.

La descripción de la base de datos que estamos utilizando se ve así:

Nombre de la columna	Tipo de dato	Descripción	Unidades
OID	Numérica	Identificador único del registro	-
Cycle OID	Numérica	Identificador único del ciclo	-
Date	Fecha	Fecha del ciclo	-
Month	Numérica	Mes del ciclo	1–12
Year	Numérica	Año del ciclo	-
Secondary Machine Name	Categórica	Nombre de la máquina secundaria (pala) (Siempre exc)	-
Secondary Machine Class Name	Categórica	Clase de la máquina secundaria	-
Material Name	Categórica	Nombre del material movido	-
Material Group	Categórica	Grupo del material (Ore/Waste)	-
Cycle Duration	Numérica	Duración del ciclo	Minutos
Completed Cycle Duration Min	Numérica	Duración del ciclo completo	Minutos
Loading Duration (min)	Numérica	Tiempo de carga	Minutos
Travelling Empty Duration (min)	Numérica	Tiempo de viaje vacío	Minutos
Queuing At Source Duration (min)	Numérica	Tiempo de espera en la fuente	Minutos
Spotting At Source Duration (min)	Numérica	Tiempo de alineación en la fuente	Minutos
Wait For Load Duration (min)	Numérica	Tiempo de espera para cargar	Minutos
Payload_Q	Numérica	Toneladas transportadas en el ciclo	Toneladas
Completed Cycle Duration	Numérica	Duración total del ciclo	Minutos

b) Pre-procesamiento y limpieza

- Carga y visualización del dataset original:

Se utilizó un archivo Excel con información detallada de los ciclos de acarreo realizados por camiones en una operación minera. Este contenía 97,718 registros con múltiples columnas.

- Filtrado de la muestra:

Se filtraron únicamente los registros correspondientes a una pala específica y a materiales clasificados como Ore o Waste. Esto se realizó para tener un conjunto de datos homogéneo en cuanto a equipo y tipo de material, y eliminar posibles lecturas erróneas del sistema.

- Creación de variable objetivo:

Se calculó una nueva columna denominada Productividad_avg, que representa la productividad puntual de cada ciclo en toneladas por hora. Esta se obtuvo mediante la

fórmula:
$$\frac{\text{Payload(Ton)}}{\text{Completed Cycle Duration (min)}} * 60$$

- Tratamiento de valores nulos:

Se eliminaron las filas con valores nulos en las columnas clave para el análisis predictivo (Payload_Q, Completed Cycle Duration y Productividad_avg). Esta depuración redujo el dataset a aproximadamente 78,904 registros limpios.

- Eliminación de outliers:

Se identificaron valores extremos (outliers) en la variable de productividad utilizando el método del rango intercuartil. (IQR). Los ciclos con valores muy altos o bajos en productividad fueron removidos para evitar que afectaran la performance de los modelos.

c) Análisis descriptivo

A partir de los 97,718 ciclos registrados inicialmente, se filtraron aquellos que cumplían con los criterios de calidad, lo que resultó en un subconjunto de aproximadamente 78,904 ciclos luego de limpieza y eliminación de valores atípicos.

Variables principales exploradas:

Payload_Q: peso del material transportado por ciclo (en toneladas).

Completed Cycle Duration: duración completa del ciclo en minutos.

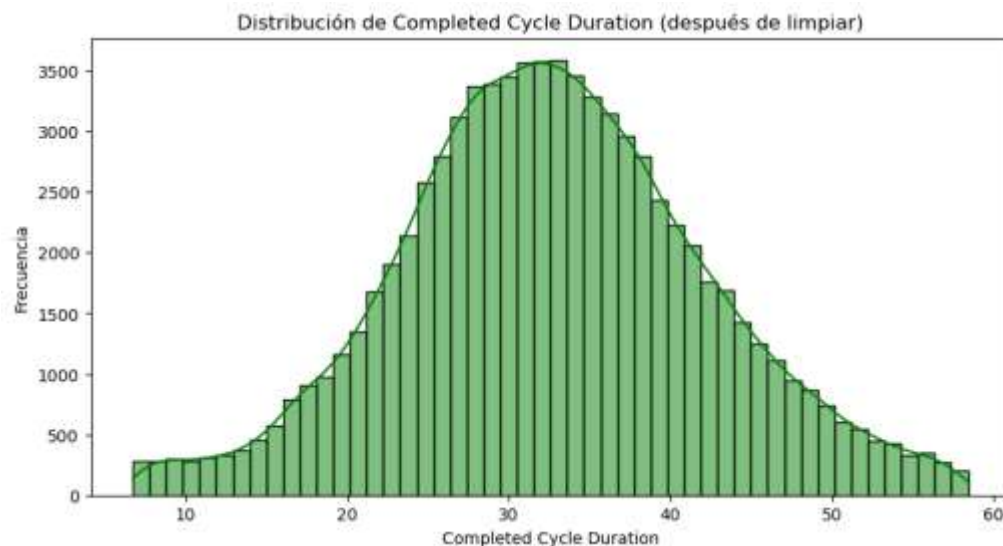
Productividad_avg: toneladas transportadas por hora en cada ciclo.

Para evaluar la manera en la que se comportaban los datos se utilizó la variable de cycle duration y productividad.

Las estadísticas descriptivas de productividad arrojaban lo siguiente:

```
Estadísticas de Productividad (ton/hora):  
count      82212.000000  
mean       808.859059  
std        7232.648139  
min         0.000000  
25%        515.416097  
50%        632.677810  
75%        777.815584  
max       999000.000000  
Name: Productividad_avg, dtype: float64
```

Y para visualizar la distribución de datos de cycle duration sacamos la gráfica de distribución normal:



d) Selección de variables

La selección de variables fue realizada considerando aquellas variables que tienen un mayor impacto en la variable objetivo del modelo: la productividad del camión medida en toneladas por hora (Productividad_avg).

A partir del análisis previo, se identificaron las siguientes variables como predictoras clave:

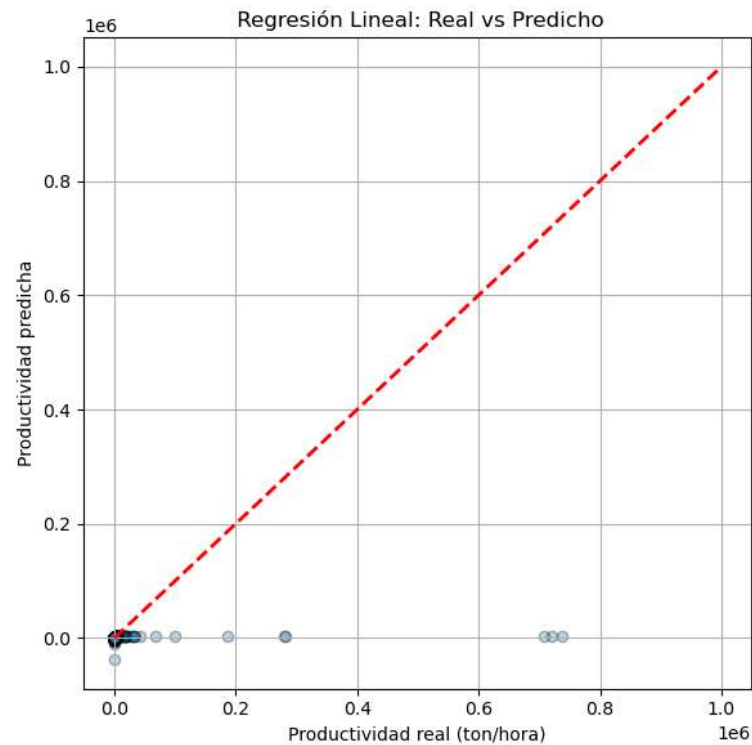
Payload_Q: cantidad de material transportado por ciclo (toneladas).

Completed Cycle Duration: duración total del ciclo completo de acarreo (en minutos).

e) Selección de Modelos

Se aplicaron tres modelos de regresión:

- Regresión Lineal: Para la regresión lineal se observó baja capacidad predictiva.

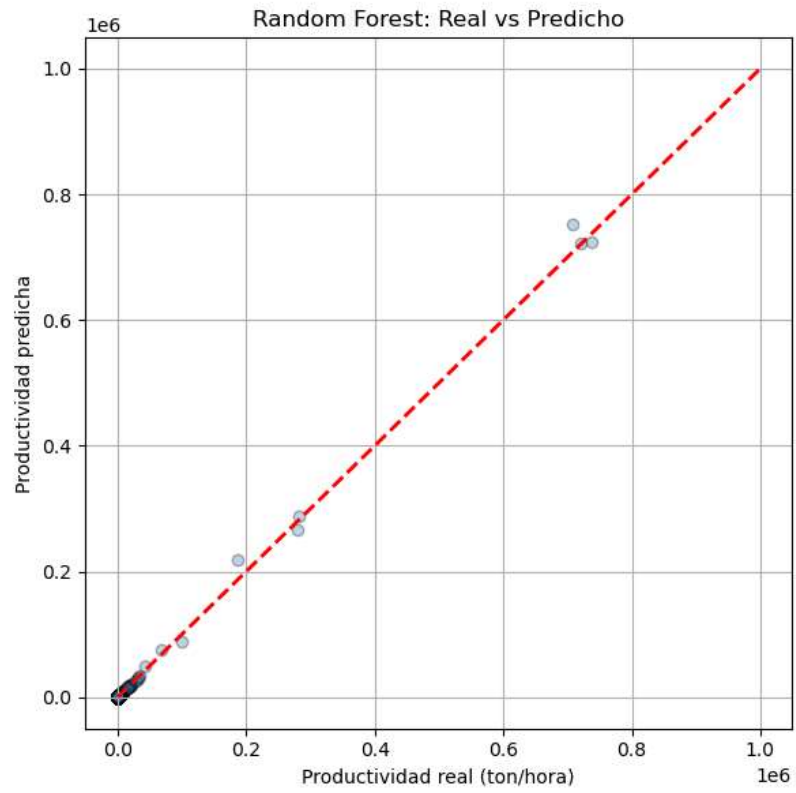


R^2 : 0.006

RMSE: 10371.75

MAE: 536.47

- Regresión Random Forest: Esta sin embargo tenía buena capacidad predictiva pero susceptible a sobreajuste.

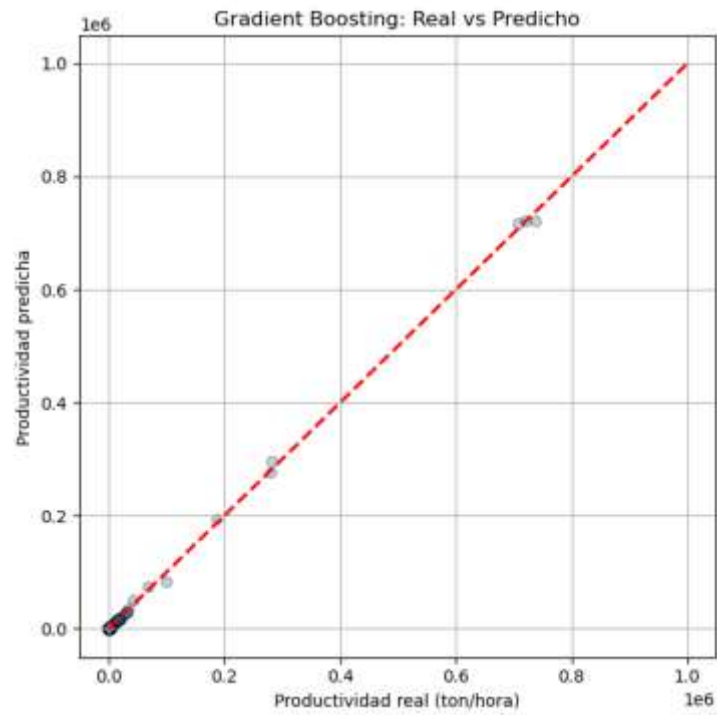


R^2 : 0.998

RMSE: 461.23

MAE: 9.99

- Regresión Gradient Boosting: el modelo que más se ajusta y menor riesgo.



R^2 : 0.999

RMSE: 250.81

MAE: 21.05

CONCLUSIONES

A lo largo del desarrollo del proyecto pude analizar en profundidad los factores que influyen en la productividad de los camiones en operaciones mineras, usando como variable objetivo las toneladas por hora (ton/hora). Al explorar más de 75 mil registros de ciclos de acarreo reales, entendí cómo variables como el payload y la duración total del ciclo impactan directamente en el rendimiento del equipo.

Durante el análisis, fue evidente que la relación entre estas variables no siempre es lineal. Por ejemplo, un ciclo más largo no necesariamente implica una menor productividad si el camión transporta una mayor cantidad de material. Esta fue una de las cosas que más me sorprendió y me llevó a entender la importancia de considerar varias variables juntas y no de forma aislada.

En cuanto a los modelos, probé diferentes enfoques, desde regresión lineal hasta modelos más complejos como Random Forest y Gradient Boosting. Fue interesante ver cómo los modelos basados en árboles se ajustaron mucho mejor a los datos, logrando métricas de error bastante bajas. Esto refuerza la idea de que para este tipo de datos operativos, con posibles relaciones no lineales, los modelos basados en árboles tienen un mejor desempeño.

Este tipo de análisis tiene el potencial de mejorar la eficiencia y productividad de manera concreta en el campo.

RECOMENDACIONES Y FUTUROS ESTUDIOS

- Como resultado de este análisis, se recomienda a futuras investigaciones en el área de productividad minera considerar el enriquecimiento del dataset con más variables operativas y contextuales, como condiciones climáticas, topografía de las rutas de acarreo, turnos de operación o eventos atípicos (por ejemplo, mantenimientos o incidentes).
- Sería útil también probar con otros enfoques como análisis de series temporales para poder hacer predicciones más allá de solo generales.
- Más información por meses y no solo 90 días.

BIBLIOGRAFÍA

Hartman, H. L., & Mutmanský, J. M. (2002). Introductory Mining Engineering. John Wiley & Sons.

Elbrond, J. (1994). Open Pit Mine Planning and Design. CRC Press.

Osanloo, M., & Ataei, M. (2004). Truck-Shovel Dispatching Model for Open Pit Mines. Journal of Mining Science.

ANEXOS

- Link de Github: https://github.com/lizperezd/miningtrucks_productivity