CLASIFICACIÓN DE RECURSOS MINERALES MEDIANTE MACHINE LEARNING; UN ENFOQUE SEMI AUTOMÁTICO

Se propone una metodología de clasificación de recursos minerales utilizando algoritmos de Machine Learning, con el objetivo de respaldar el trabajo del geólogo competente, reducir costos temporales y disminuir la subjetividad en el proceso de clasificación.

Heber Hernández Guerra
Académico Ingeniería Civil en Minas

Cristian Sánchez López
Director de Carrera Ingeniería Civil en Minas

Dependiendo del nivel de confianza geológica, los recursos minerales de un depósito minero se clasifican en categorías de medidos, indicados e inferidos, derivado de los resultados de la estimación, calidad y cantidad de datos disponibles en un modelo de bloques minero. Aunque existen múltiples criterios aceptados para lograr una de las categorías de clasificación, actualmente no existe una estandarización que determine el más adecuado, y los umbrales de corte a utilizar. Esta situación conduce a la toma de una decisión subjetiva a cargo de la persona competente (QP), contradiciendo la necesidad de objetividad y dejandola de decisión principalmente a juicio experto. Este artículo, presenta una metodología innovadora para la clasificación de recursos minerales mediante el empleo de algoritmos de aprendizaje de máquinas. El propósito principal de esta metodología es respaldar la labor del QP, reducir los costos temporales, disminuir la subjetividad y simplificar la reproducción y auditabilidad del proceso, alineándose con los principios fundamentales de la clasificación.

La metodología propuesta consta de tres etapas secuenciales. En primer lugar, el QP selecciona las variables que definirán la clasificación, grabadas bloque a bloque, según su criterio. Estas variables pueden ser de naturaleza geométrica, geoestadística, geológica o de cualquier tipo, tanto cuantitativas como cualitativas. Posteriormente, a través de aprendizaje no supervisado y utilizando agrupamiento automático mixto (k-prototipos), se lleva a cabo un agrupamiento multivariable por similitud mediante la distancia Gower. Finalmente, mediante aprendizaje supervisado con una red neuronal multicapa, se suavizan los bloques según la categoría de recursos.

La metodología se aplica en un caso 3D real para un depósito hidrotermal de oro de alta sulfuración. Los resultados obtenidos reflejan los principios fundamentales de la clasificación: transparencia, objetividad, reproducibilidad y auditabilidad basadas en principios cuantificables. Además, la metodología se muestra coherente con los criterios empleados y es comparable con la metodología tradicional.

Introducción

Un recurso mineral se define como una concentración u ocurrencia de material sólido de interés económico en la corteza terrestre, con la forma, ley, calidad y cantidad que ofrece perspectivas razonables para su eventual extracción económica. La ubicación, cantidad, ley o calidad, continuidad y otras características geológicas de un recurso mineral son conocidas, estimadas o interpretadas a partir de evidencia y conocimiento geológico específico. El recurso mineral se clasifica en orden ascendente de confianza geológica, en las categorías de Inferido, Indicado y Medido. Esta subdivisión refleja la confianza en la cantidad y calidad estimada de la mineralización y depende de la fiabilidad de las interpretaciones geológicas. La clasificación es esencial para empresas mineras, inversionistas e instituciones financieras, ya que las decisiones de inversión suelen basarse en la ley, tonelaje y confianza asignada a los depósitos. La progresión hacia la presentación de reservas minerales, correspondiente a la parte económica del recurso, requiere al menos la clasificación de recursos indicados y/o medidos, reconociendo la asignación a una categoría indicada como paso necesario para avanzar hacia la viabilidad del proyecto.

Aunque los códigos internacionales para reportar recursos y reservas mineras requieren una clasificación sólida, los criterios para definir estas clasificaciones son numerosos; geométricos, geoestadísticos y combinaciones de ambos, entendiendo que por cada uno de estos enfoques existen decenas de variaciones. Diferentes criterios generan variación en ley mineral, cantidad de metal y tonelaje asociado a cada categoría. Todos los criterios son ocupados estableciendo umbrales de manera manual, inclusive cuando son combinados en paralelo. Muchas veces los resultados son espacialmente incoherentes y no se respeta la continuidad geológica intrínseca, por lo que requieren de un procesamiento posterior de suavizado manual.

Este artículo no tiene como objetivo determinar el criterio más adecuado entre los disponibles, ya que todos son validables y cuestionables al mismo tiempo. Más bien, se propone que la selección de criterios sea tarea del QP, quien puede combinarlos de manera coherente según el conocimiento del depósito mineral. Este estudio propone automatizar el proceso de clasificación según los criterios seleccionados por el QP (etapa 1), evitando el establecer umbrales de forma manual, para posteriormente realizar un suavizamiento de los bloques en dos etapas posteriores, las que se describen a continuación:

- Agrupamiento multivariable mixto bloque a bloque a través de k-prototipos, a través de un algoritmo de aprendizaje de máquinas no supervisado sin limitaciones en la cantidad de criterios de entrada, ni en el tipo de dato, ya sea cuantitativo o cualitativo.
- Suavizamiento de los bloques mediante una red neuronal multicapa, para mitigar el efecto de "spotted dog" propio del uso de criterios geométricos y geoestadísticos bloque a bloque.

Con esta nueva metodología, la clasificación permite principios de transparencia y objetividad. Además, garantiza que el método de clasificación sea reproducible y auditable basado en principios cuantificables, los resultados de las categorías de recursos sean contiguos y reflejen fielmente su relación con los criterios de entrada. Finalmente, esta metodología impacta directamente en la reducción del costo temporal e intervención humana.

Desarrollo

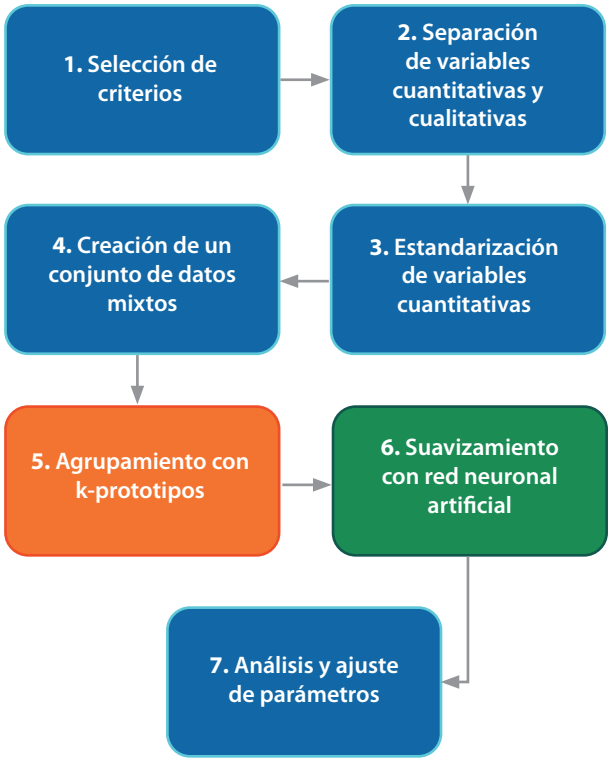

Figura 1

Flujo de trabajo para la clasificación de recursos minerales mediante la metodología propuesta.

El QP selecciona los criterios a utilizar para la clasificación de los recursos minerales, pudiendo ser estos geométricos, geoestadísticos, geológicos, o de cualquier índole que este considere significativos para el proceso. Estas variables deberán estar disponibles bloque a bloque en el modelo discretizado del depósito mineral. Luego se separarán aquellas variables cuantitativas de las cualitativas, para que estas primeras se sometan a un proceso de transformación de potencia realizada con el método Yeo-Johnson. Este método estabiliza la varianza y disminuye la asimetría. Las variables transformadas cuantitativas se unirán en una nueva tabla de datos junto con las variables cualitativas. Cabe señalar que este proceso no es excluyente, pudiendo utilizarse únicamente variables cuantitativas si así lo decide el QP. Cuando esto último ocurre, k-prototipos opera como k-medias. Posteriormente se lleva a cabo el agrupamiento, para lo cual se requiere un parámetro que es el número de grupos, sin embargo, los recursos minerales siempre se clasificaran en tres categorías, por lo que este parámetro queda definido como una constante y el proceso pasa a ser automático.

El algoritmo k-prototypes permite el uso de datos tanto numéricos como categóricos. El algoritmo principal consiste en calcular la distancia entre objetos de datos y centros de grupo o prototipos (Q). El método finaliza cuando el centro de grupo actualizado coincide con el centro de grupo de la iteración anterior.

Dado un conjunto de datos de n objetos, donde m_r son los atributos numéricos y m_c los categóricos, el objetivo de k-prototypes es encontrar k grupos donde se minimice la siguiente función objetivo:

$$\sum_{l=1}^{k} \sum_{i=1}^{n} p_{il}\, d(x_i, Q_l),$$

donde p_{il} es un elemento de la matriz de partición $p_n \cdot k$, satisfaciendo $0 \leq p_{il} \leq 1$ y $\sum_{l=1}^{k} y_{il} = 1$. P es una partición rígida si p_{il} es una variable binaria ($p_{il} \in \{0, 1\}$) que indica la pertenencia de los datos x_i en el cluster l; de lo contrario, P es una partición difusa. Q_l es el centro o prototipo del clúster l y $d(x_i, Q_l)$ es la medida de distancia definida de la siguiente manera:

$$d(x_i,Q_l)= \sum_{j=1}^{m_r} (x_{ij}^{r} - q_{lj}^{r})^2 + \gamma_l \sum_{j=1}^{m_c} \delta(x_{ij}^{c} - q_{lj}^{c}),$$

x_{ij}^{r} representa los valores de los atributos numéricos y x_{ij}^{r} los valores de los atributos categóricos para cada objeto de datos i; q_{lj}^{r} es la media del j^{th} atributo numérico en un clúster l, and q_{lj}^{c} es la moda de j^{th} atributo categórico en el clúster l; γ_l es un peso para atributos categóricos en un clúster l. La función δ, que está definido para atributos categóricos, es:

$$\delta(x_{ij}^{c}, q_{lj}^{c}) = \{0, x_{ij}^{c} \neq q_{lj,}^{c} \quad 1, \quad x_{ij}^{c} = q_{lj}^{c}.$$

La ecuación de la función de costo para tipos de datos mixtos (numéricos y categóricos) es la siguiente:

$$Cost_l = \sum_{l=1}^{k} u_{il} \sum_{j=1}^{m_r} (x_{ij}^{r} - z_{ij}^{r})^2 + \gamma_l \sum_{j=1}^{m_c} u_{il} \sum_{j=1}^{m_c} \delta(x_{ij}^{c}, z_{ij}^{c}),$$

Simplificado de forma práctica como:

$$Cost_l = Cost_l^{r} + Cost_l^{c},$$

donde $Cost_l^{r}$ denota el costo total de todas las variables numéricas para todos los objetos dentro del cluster l. $Cost_l^{r}$ se minimiza siempre y cuando z_{ij} se calcula utilizando la siguiente ecuación:

$$z_{ij} = \frac{1}{n_l} \sum_{i=1}^{n} u_{il} \cdot x_{ij} \; ; \; para \; j = 1,2,...,m,$$

donde $n_l = \sum_i^n u_{il} \cdot x_{ij}$ es el número de objetos dentro del clúster l. Para variables categóricas, Cj es un conjunto de valores únicos en cada variable categórica j, y $p(l)$ es la probabilidad de que Cj está dentro del clúster l. Por lo tanto, el costo se puede reescribir de la siguiente manera.

$$Cost_i^{n} = \gamma_l \sum_{j=1}^{m_c} n_l (1-p(l)),$$

donde n_l es el objeto dentro del clúster. Asumiendo el uso de criterios geométricos y geoestadísticos en el agrupamiento, los resultados son suavizados por una red neuronal, utilizando como características de entrada las coordenadas espaciales, y como variable objetivo la categoría previamente recomendada por k-prototipos; medido, indicado o inferido.

Una red neuronal artificial es un sistema compuesto por múltiples nodos de procesamiento simples que operan en paralelo y cuya función está determinada por la estructura de la red y el peso de las conexiones, donde el procesamiento se realiza en cada uno de los nodos. Las redes MLP pertenecen a la arquitectura de alimentación múltiple en capas, cuyo entrenamiento se realiza con un proceso supervisado. El término feedforward se refiere al flujo de información a través de la red de manera unidireccional, desde la capa de entrada hasta la capa de salida. En general, una NN se puede dividir en tres partes: capa de entrada, capas ocultas (h=1,2,...,n) y capa de salida. La capa de entrada recibe información del entorno externo, que suele normalizarse para mejorar la precisión numérica en las operaciones matemáticas de la red. Las capas ocultas están compuestas por neuronas que se encargan de extraer patrones asociados con el proceso o sistema analizado. Estas capas realizan la mayor parte del procesamiento interno de la red. La capa de salida se encarga de generar y presentar las salidas finales de la red. Estas salidas se derivan del procesamiento realizado por las neuronas en las capas anteriores. Cada neurona en las capas ocultas se encarga de recibir información de la capa anterior, procesarla y propagarla hacia adelante para llegar a la capa de salida. Lo que sucede en cada neurona se puede explicar a través del principio del perceptrón.

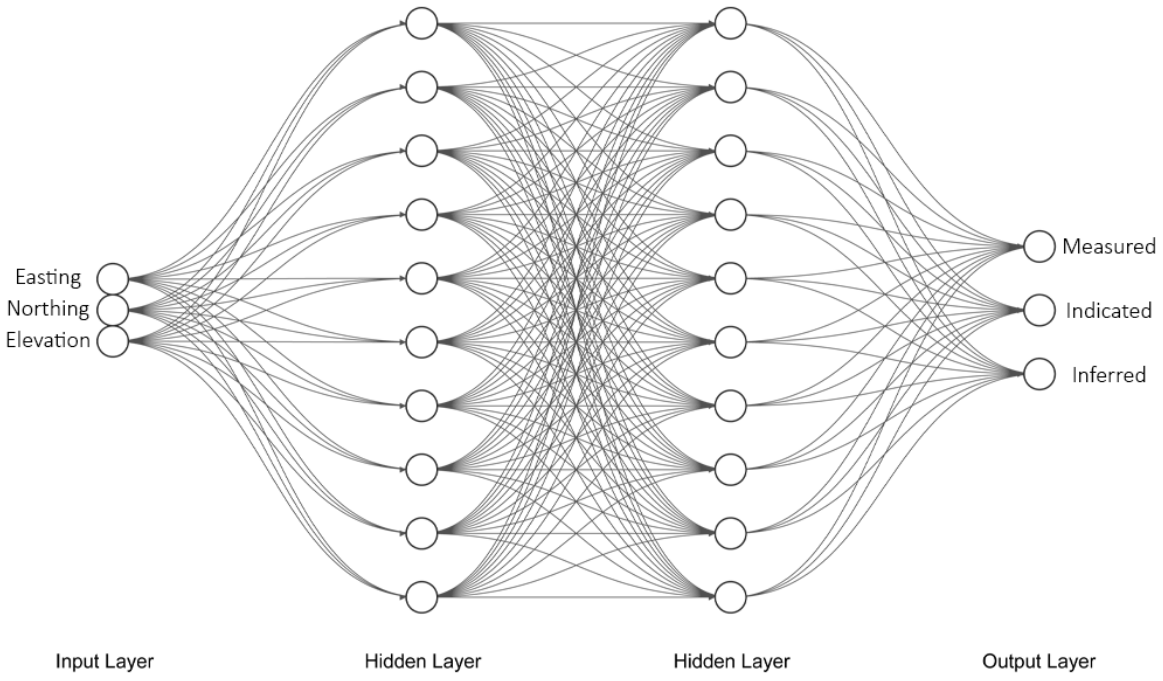

Figura 2

Arquitectura MLP propuesta para el suavizado de bloques.

$$\hat{y}=f(\sum_j x_j w_j+\theta),$$

donde los x_j son las entradas a la unidad, los w_j son los pesos, θ es el término de sesgo, f es la función de activación no lineal, y $\hat{y}$ es la activación de la unidad. Extendiendo este concepto a múltiples capas, se calculan las salidas reales de la red:

$$\hat{y}_k(p) = f(\sum_{i=1}^{m} x_{jk}(p) \cdot w_{jk} + \theta_k),$$

Donde m es el número de entradas para la neurona k desde la capa de salida, y f es la función de activación. La función de activación de cada neurona también es un hiperparámetro de la red neuronal. En este caso, se utiliza la Unidad Lineal Rectificada (ReLU) para las capas ocultas y SoftMax para la capa de salida.

$$f(x)=max\ (0,x)\{\ 0\ for\ x<0\ x\ for\ x\geq 0$$

Donde x, es el dato de entrada o vector de características $\sum_{i=1}^{m} x_{jk}(p)\cdot w_{jk}+\theta_k$.

$$f(x)_i = \frac{e^{xi}}{1+\sum_{j=1}^{c} e^{xj}}, i=1,...,C$$

El modelo es entrenado por retroprogramación con el optimizador ADAM. Finalmente, la clasificación de los recursos es analizada a nivel visual y a nivel estadístico con respecto a las variables de entrada. Si el QP está conforme se da por cerrado el proceso, y en caso de realizar ajustes, primeramente, se deberá revisar la idónea selección de variables de entrada, y luego el suavizado de bloques a través de hiperparametros del modelo de red neuronal (número de nodos en las capas ocultas específicamente).

Análisis

Se dispone de datos derivados de la estimación realizada en un depósito de oro hidrotermal de alta sulfuración situado en el sur de Perú. Esta estimación se presenta de manera resumida en un modelo discretizado de bloques 3d de tamaño 10 x 10 x 10 metros cúbicos, del cual se ha seleccionado trabajar con las variables, ley mineral de Au estimado con KO, KV, AD, NS y GC. Asimismo, se cuenta con la información utilizada durante dicho proceso de estimación en un único dominio, la cual se obtuvo a partir de 131 perforaciones de diamantina y 37 zanjas de exploración.

Con motivos ilustrativos, solo se visualiza una sección Este-Elevación (X-Z) en el centro de la coordenada Norte (Y). En la figura 4 se pueden ver las cuatro variables seleccionadas para este caso, las cuales son las mismas que en el caso sintético 2D. Se observa que entre los 450 a 650 metros Este (m), cercano a la superficie, existe más información y la calidad de esta es alta debido a que proviene de sondajes diamantinos, por lo que existe una mayor confianza geológica. La KV captura la forma y dirección de los sondajes en dicha sección, lo que evidencia un alcance reducido del variograma, muy típico en cuanto a la variabilidad de metales preciosos como el oro. Sin embargo, no es alcance de este artículo profundizar sobre los parámetros del variograma y el plan de estimación sobre los datos disponibles.

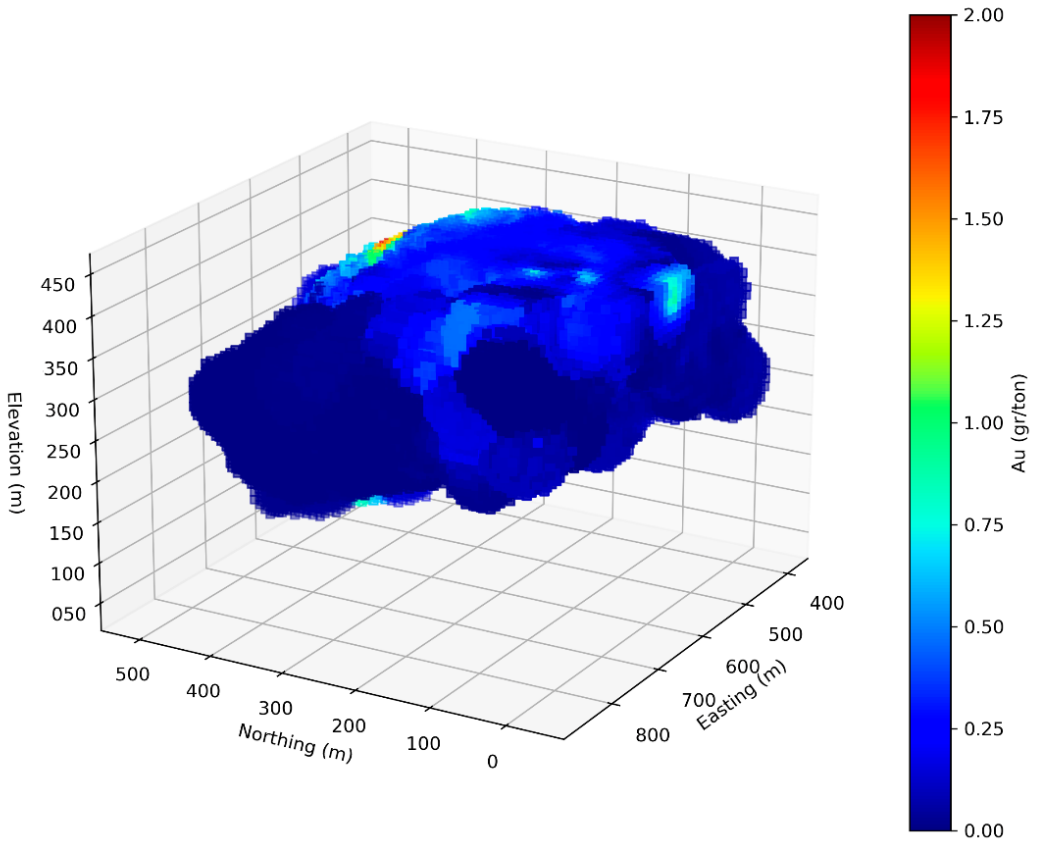

Figura 3

Bloques de Au estimados por el método de OK.

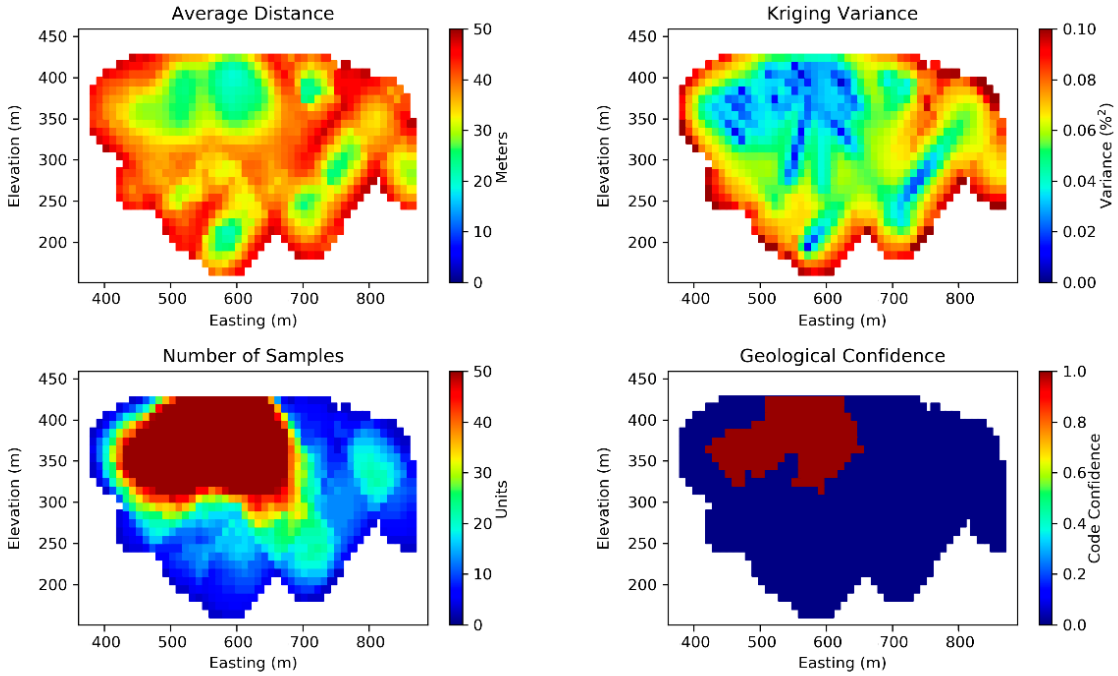

Figura 4

Variables de entrada al proceso de agrupamiento automático mixto.

KV, AD y NS son transformadas, mientras que GC es ingresada como texto a k-prototipos. El resultado es bastante coherente ya que los recursos medidos se limitan a la zona con más información, sin embargo, también se clasifican con la máxima confianza bloques que circundan sondajes aislados. En este artículo se propone resolver el problema de spotted dog, suavizando los bloques con un modelo de red neuronal de dos capas, con funciones de activación ReLu y función Softmax en las tres neuronas de salida. El proceso modifica las categorías en aquellos bloques que no se encontraban contiguos.

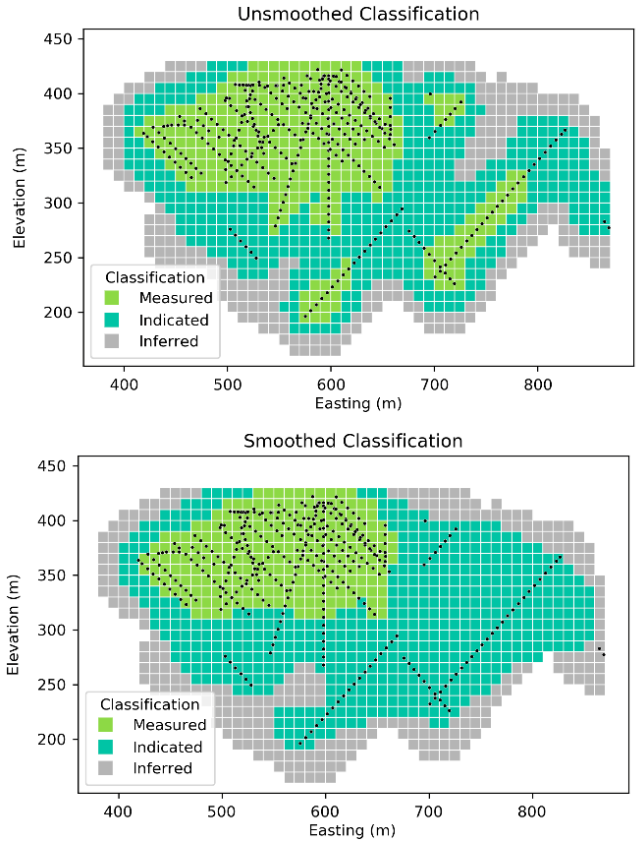

Figura 5

Recursos clasificados sin suavizar (izquierda), y suavizados (derecha).

Tabla 1. Resultados de la clasificación por tonelaje, ley y metal.

	Tonelaje	Ley (gt/ton)	Metal (ton)
Medidos	19,702,500	0.22	4.33
Indicados	67,382,500	0.11	7.41
Inferidos	33,447,500	0.10	3.34

Las cuatro variables seleccionadas para este caso tienen una clara diferenciación entre categorías de recursos, siendo los medidos los que poseen menores VK, AD y mayores NS. El modelo explícito de confianza geológica en un 100% se encuentra dentro de los recursos medidos. Estos resultados a diferencia de un sistema tradicional de reglas rígidas, permite que bloques con valores similares en KV, puedan estar clasificados en distintas categorías, ya que la clasificación considera otros factores adicionales paralelamente y al mismo tiempo busca respetar la continuidad espacial.

Conclusiones

La implementación del algoritmo de agrupamiento k-prototipos ha demostrado que es capaz de definir categorías de recursos combinando variables numéricas y categóricas en un proceso mixto automático, dado a que el único parámetro requerido es una constante igual a tres para el problema de clasificación. Es crucial, sin embargo, que el QP realice una selección adecuada de las características o criterios de entrada que se usen en el agrupamiento. Las que sean características numéricas cuantitativas deben presentar correlación ya sea positiva o negativa. La etapa siguiente de suavizamiento a través de la red neuronal mitiga el efecto de spotted dog, típico en el uso de criterios geométricos y geoestadísticos, y permite al QP realizar cambios en los hiper parámetros del modelo según sea necesario posterior analizar los resultados.

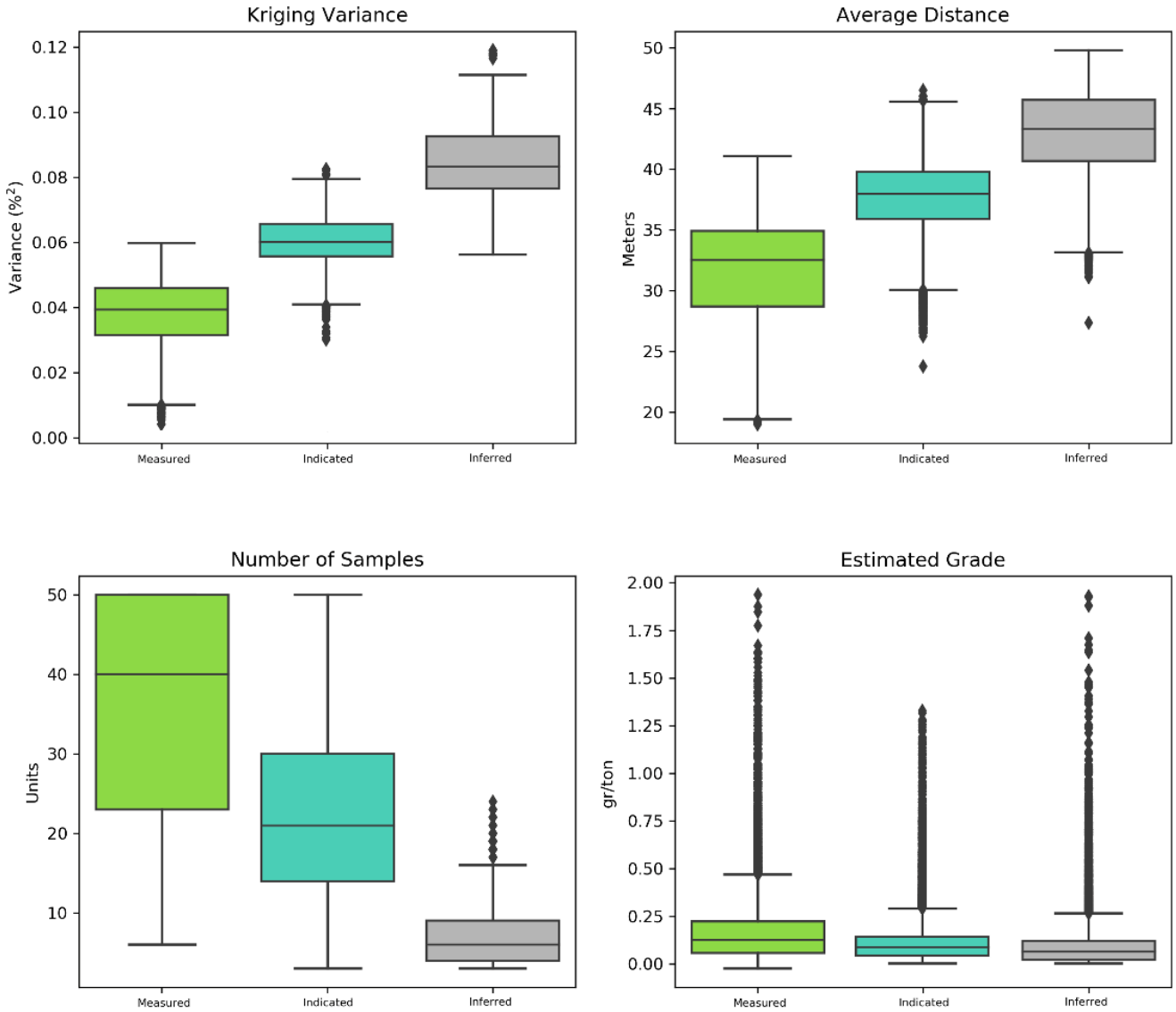

Figura 6

Diagramas de cajas y bigotes de cada característica de entrada por categorías.

La metodología propuesta propone una clasificación transparente, clara, reproducible y auditable de forma sencilla en todas sus etapas, además de favorecer que las categorías de recursos sean contiguas y fundamenten el nivel de confianza a través de los criterios seleccionados por el QP. Esto último añade versatilidad, dado que es posible incorporar cualquier tipo de criterio, y que estos trabajen en simultáneo.

En última instancia, la metodología propuesta representa una significativa reducción en el costo temporal, evitando la determinación manual de umbrales, la ponderación manual de los criterios en un modelo lineal combinado, y minimizando la intervención humana en el proceso de suavizado.

Referencias

[1] Taghvaeenezhad et al. (2020). Quantifying the criteria for classification of mineral resources and reserves through the estimation of block model uncertainty using geostatistical methods: a case study of Khoshoumi Uranium deposit in Yazd, Iran. Geosystem Engineering, 23(4), 216-225.

[2] Madani, N. (2020). Mineral Resource Classification Based on Uncertainty Measures in Geological Domains. En E. Topal (Ed.), Proceedings of the 28th International Symposium on Mine Planning and Equipment Selection - MPES 2019. MPES 2019. Springer Series in Geomechanics and Geoengineering. Springer, Cham.

[3] Cevik et al. (2021). On the Use of Machine Learning for Mineral Resource Classification. Mining, Metallurgy & Exploration, 38(5), 2055–2073.

[4] Parker, H. M., & Dohm, C. E. (2014). Evolution of mineral resource classification from 1980 to 2014 and current best practice. In FINEX 2014 Conference.

[5] Abzalov, M. (2016). Methodology of the Mineral Resource Classification. En Applied Mining Geology. Modern Approaches in Solid Earth Sciences, vol 12. Springer, Cham.