
**PREDICCIÓN Y DIAGNÓSTICO EN SUELOS CONTAMINADOS POR DAM
USANDO MACHINE LEARNING**

**JOAQUÍN EDUARDO RETAMAL CEPEDA
INGENIERO CIVIL EN MINAS**

RESUMEN

El suelo es un recurso natural que juega un papel importante en la sostenibilidad de los ecosistemas, ya que sirve de soporte para todos los seres vivos. Además, de suministrarles el agua y los nutrientes que necesitan para su completo desarrollo. Una contaminación de este podría generar una alteración desfavorable, pudiendo disminuir la calidad de las funciones que desempeña y eventualmente suponer un riesgo a la salud humana o medio ambiente. En Chile existen varias actividades económicas con potencial de contaminar el suelo; actividad forestal, agrícola y minería. Particularmente, la actividad minera puede desencadenar una de las problemáticas ambientales más relevantes, el drenaje ácido minero (DAM) que se genera a partir de la oxidación de sulfuros como la pirita, en presencia de oxígeno atmosférico y agua, produciéndose agua ácida cargada con sulfatos, metales y metaloides que pueden alcanzar altas concentraciones, generando un riesgo para el medio ambiente. Por lo tanto, esta memoria tiene como objetivo predecir y diagnosticar la contaminación de suelo causada por DAM. Para esto, se requiere la caracterización del emplazamiento contaminado y posteriormente, efectuar la técnica de estimación que se considere adecuada para determinar concentración de un elemento contaminante, tal como: métodos deterministas o geoestadístico. Aunque, dentro de las motivaciones de esta memoria, se busca considerar el machine learning como una nueva alternativa de estimación mediante el uso de redes neuronales. Para esto, se realiza un análisis de datos a través de métodos estadísticos para averiguar comportamiento y relaciones que existen entre variables. Posteriormente, se divide la base de datos: conjunto de entrenamiento, aquel que proporciona datos a las redes que le permiten aprender y conjunto de prueba, aquellos que serán utilizados para evaluar capacidad predictiva del modelo. Finalmente, se compara modelo machine learning con modelo geoestadístico por medio de matriz de confusión. Los resultados obtenidos

en esta memoria, muestran que cuando modelo geoestadístico y machine learning trabajan con variables continuas existen mínimas diferencias. Sin embargo, cuando modelos trabajan con variables categóricas, el método de kriging ordinario obtiene un 10% de zonas de incertidumbre. En cambio, las redes neuronales predicen entre un 3% y 7% zonas de incertidumbre. Además, de obtener mayor porcentaje de acierto en los puntos altos de contaminación, entregando como resultado 6.7 hectáreas contaminadas, lo que equivale al 20% del área de estudio.

ABSTRACT

Soil is a natural resource that plays an important role in the sustainability of ecosystems, since it serves as a support for all living things. In addition, it provides water and nutrients needed for their development. Soil contamination could generate an unfavorable alteration, being able to reduce the quality of the functions it performs and eventually being a risk to human health or the environment. In Chile, there are several economic activities with the potential to contaminate the soil; forestry, agricultural and mining activity. Particularly, mining activity can trigger one of the most relevant environmental problems, acid mine drainage (DAM) that is generated from the oxidation of sulfides such as pyrite, in the presence of oxygen and water, producing acidic water loaded with sulfates, metals and metalloids that can reach high concentrations, creating a risk for the environment. Therefore, this report is intended to predict and diagnose soil contamination caused by DAM. For this, the characterization of the contaminated site is required and subsequently, perform the estimation technique considered appropriate to determine the concentration of a contaminating element, such as: deterministic or geostatistical methods. Although, within the motivations of this report, I seek to consider machine learning as a new alternative of estimation through the use of neural networks. For this, a data analysis is performed through statistical methods to find out the behavior and relationships that exist between variables. Subsequently, the database is divided: training set, the one that provides data to the networks that allow it to learn, and the test set, those that will be used to evaluate the predictive capacity of the model. Finally, a machine learning model is compared with a geostatistical model using a confusion matrix. The results obtained in this report show that when geostatistical model and machine learning work with continuous variables there are minimal differences. However, when models work with categorical variables, the ordinary kriging method gets 10% uncertainty zones. In contrast, neural networks predict between 3% and 7% uncertainty zones. In addition, neural networks to obtain a higher percentage of success in the high

points of contamination, giving as a result 6.7 contaminated hectares, which is equivalent to 20% of the study area.