

Bullejos, Manuela;
Cabezas, Davidb;
Martín-Martín, Manuelc
Alcalá, Francisco Javier

Abstract

In a previous paper, the authors implemented a machine learning k-nearest neighbors (KNN) algorithm and Python libraries to create two 3D interactive models of the stratigraphic architecture of the Quaternary onshore Llobregat River Delta (NE Spain) for groundwater exploration purposes. The main limitation of this previous paper was its lack of routines for evaluating the confidence of the 3D models. Building from the previous paper, this paper refines the programming code and introduces an additional algorithm to evaluate the confidence of the KNN predictions. A variant of the Similarity Ratio method was used to quantify the KNN prediction confidence. This variant used weights that were inversely proportional to the distance between each grain-size class and the inferred point to work out a value that played the role of similarity. While the KNN algorithm and Python libraries demonstrated their efficacy for obtaining 3D models of the stratigraphic arrangement of sedimentary porous media, the KNN prediction confidence verified the certainty of the 3D models. In the Llobregat River Delta, the KNN prediction confidence at each prospecting depth was a function of the available data density at that depth. As expected, the KNN prediction confidence decreased according to the decreasing data density at lower depths. The obtained average-weighted confidence was in the 0.44–0.53 range for gravel bodies at prospecting depths in the 12.7–72.4 m b.s.l. range and was in the 0.42–0.55 range for coarse sand bodies at prospecting depths in the 4.6–83.9 m b.s.l. range. In a couple of cases, spurious average-weighted confidences of 0.29 in one gravel body and 0.30 in one coarse sand body were obtained. These figures were interpreted as the result of the quite different weights of neighbors from different grain-size classes at short distances. The KNN algorithm confidence has proven its suitability for identifying these anomalous results in the supposedly well-depurated grain-size database used in this study. The introduced KNN algorithm confidence quantifies the reliability of the 3D interactive models, which is a necessary stage to make decisions in economic and environmental geology. In the Llobregat River Delta, this quantification clearly improves groundwater exploration predictability. © 2023 by the authors.

Author keywords

3D stratigraphic architecture; confidence degree; data classification; KNN algorithm; Llobregat River Delta; python libraries