Signals in the Soil: Empowering Data-Driven Simple Models with Feature-Rich and Data's Lens for Forecasting of Landslide Displacement in the Daunia (Bovino Region), Italy

Milad Sabaghi
Department of Earth and Environmental Sciences, University of Bari Aldo Moro, Bari, Italy
Piernicola Lollino
Department of Earth and Environmental Sciences, University of Bari Aldo Moro, Bari, Italy
Mario Parise
Department of Earth and Environmental Sciences, University of Bari Aldo Moro, Bari, Italy

Abstract:
A clear understanding of the displacement trends of slow landslides, both in the past and at present, can be essential for a reliable prediction of the landslide kinematical evolution in the future, in the perspective of early warning and hazard mitigation. In this context, machine learning methods have made significant advancements in the investigation of natural hazards, in general, and, specifically, of landslide processes. Such methods demonstrate superior performance by exploiting techniques like temporal lags, feature engineering, and transformations. This study examines the performance of various machine learning approaches along with several pre-processing and feature engineering techniques, in the prediction of the kinematics of slow active landslides. The Daunia area, located in the Apulia region (Southern Italy), and specifically the Bovino municipality area, is here used as a test area to verify the potentialities of two common machine learning algorithms: Linear Regression (LR) and Prophet. In particular, the efficiency of these simple algorithms shows notable improvements when feature selection techniques are implemented. The accuracy of these simple algorithms shows notable improvements before and after implementing feature selection. A key focus of this study is highlighting the efficiency of feature engineering. The results clearly demonstrate that pre-processing methods, along with temporal lags, Fourier series, and Yeo-Johnson transformations, can significantly enhance the performance of simple linear machine learning models. Also, when Linear Regression and Prophet Models are compared, the Prophet model exhibits relatively better performance. Initially, the models were tested with selected data point that included sharp changes and significant anomalies. Later on, the model with superior performance was validated using a large number of critical points. The optimal feature selection changes depending on the chosen model and data used. The evaluation