A Machine Learning Framework for Cereal Yield Forecasting Using Heterogeneous Data

Noureddine Jarray[1,2(✉)], Ali Ben Abbes[1], and Imed Riadh Farah[1]

[1] National School of Computer Science, Mannouba, Tunisia
jarraynoureddine@gmail.com
[2] Institut des Régions Arides (IRA), Medenine, Tunisia

Abstract. The combination between Machine learning (ML) and heterogeneous data offers an opportunity for agronomists in crop yield forecasting for decision support systems. ML has recently emerged as a method to support Sustainable Development Goals (SDGs) such as the SDG-2 that which aims to achieve food security. This paper presents a modularized, robust process, data-driven and ML based framework, designed to help the decision makers within the yield forecasting to accurately achieve food security. Four ML models, including eXtreme Gradient Boost (XGBoost), Random Forest (RF), Linear Regressor (LR), and Support Vector Regression (SVR) were employed to yield forecasting. Experiments were carried out on twenty provinces in Tunisia from 2002 to 2018. The results obtained showed that XGBoost slightly outperformed other ML techniques. The results of model validation, obtained from the XGBoost model showed that the Pearson correlation r, Root-Mean-Square Error (RMSE) and the Mean Absolute Error (MAE) values were 0.97, 208.492, and 105.910, respectively. This paper showed the best results and it can be used to address national food security challenges.

Keywords: Machine learning · heterogeneous data · forecasting · cereal yield · Sustainable Development Goals (SDGs)

1 Introduction

In recent decades, the advent of publicly-available heterogeneous data and the appearance of artificial intelligence (AI) played a main important role in the achievement of the SDGs [1]. They are supported by several agricultural challenges [2–6]. The objective of the National SDG-2 in to provide needy people with food to people in need [7]. The food security can be guaranteed by crop yield forecasting [8]. Wheat and barley are essential food crops. In addition, crop yield prediction is involved due to the complex relationships between crop growth and environmental variables. It is hard to analysis the crop growth process and several pertinent variables (e.g. varieties, soil management, etc.).

A. Abraham et al. (Eds.): ISDA 2022, LNNS 715, pp. 21–30, 2023.
https://doi.org/10.1007/978-3-031-35507-3_3