DEVELOPING MACHINE LEARNING MODELS FOR THE PREDICTION OF AGRONOMIC PHENOTYPES INTERESTS USING HIGH-RESOLUTION IMAGES AND GENOMIC DATA

by Kismiantini, Osval Antonio Montesinos-López, Dhoriva Urwatul Wustqa Ezra Putranda Setiawan

ABSTRACT

Genomic selection (GS) is revolutionizing plant breeding since the selection process is done with the help of statistical machine learning methods. A model is trained with a reference population and then it is used for predicting the candidate individuals available in the testing set. However, given that breeding phenotypic values are very noisy, new models must be able to integrate not only genotypic and environmental data but also high-resolution images that have been collected by breeders with advanced image technology. For this reason, this paper explores the use of generalized Poisson regression for genome-enabled prediction of count phenotypes using genomic and high-resolution images. The generalized Poisson regression model allows integrating input information of many sources like environments, genomic data, high resolution data, and interaction terms between these three sources. We found that the best prediction performance was obtained when the three sources of information were taken into account in the predictor, and those measures of high-resolution images close to the harvest day provided the best prediction performance.

Kata Kunci: *high-resolution images, genomic data, plant breeding, generalized Poisson regression, genomic selection, count data*