Integrating Data Mining with Case-Based Reasoning for Diagnosing Sorghum Anthracnose

Authors : Mariamawit T. Belete

Abstract : Cereal production and marketing are the means of livelihood for millions of households in Ethiopia. However, cereal production is constrained by technical and socio-economic factors. Among the technical factors, cereal crop diseases are the major contributing factors to the low yield. The aim of this research is to develop an integration of data mining and knowledge based system for sorghum anthracnose disease diagnosis that assists agriculture experts and development agents to make timely decisions. Anthracnose diagnosing systems gather information from Melkassa agricultural research center and attempt to score anthracnose severity scale. Empirical research is designed for data exploration, modeling, and confirmatory procedures for testing hypothesis and prediction to draw a sound conclusion. WEKA (Waikato Environment for Knowledge Analysis) was employed for the modeling. Knowledge based system has come across a variety of approaches based on the knowledge representation method; case-based reasoning (CBR) is one of the popular approaches used in knowledge-based system. CBR is a problem solving strategy that uses previous cases to solve new problems. The system utilizes hidden knowledge extracted by employing clustering algorithms, specifically K-means clustering from sampled anthracnose dataset. Clustered cases with centroid value are mapped to jCOLIBRI, and then the integrator application is created using NetBeans with JDK 8.0.2. The important part of a case based reasoning model includes case retrieval; the similarity measuring stage, reuse; which allows domain expert to transfer retrieval case solution to suit for the current case, revise; to test the solution, and retain to store the confirmed solution to the case base for future use. Evaluation of the system was done for both system performance and user acceptance. For testing the prototype, seven test cases were used. Experimental result shows that the system achieves an average precision and recall values of 70% and 83%, respectively. User acceptance testing also performed by involving five domain experts, and an average of 83% acceptance is achieved. Although the result of this study is promising, however, further study should be done an investigation on hybrid approach such as rule based reasoning, and pictorial retrieval process are recommended.

Keywords : sorghum anthracnose, data mining, case based reasoning, integration
Conference Title : ICDAAF 2023 : International Conference on Data Analytics in Agriculture and Food
Conference Location : Toronto, Canada
Conference Dates : September 18-19, 2023