Machine Learning Models for Crop Yield Prediction and a Smart Irrigation System

Marwen Elkamel and Luis Rabelo
Industrial Engineering & Management Systems Department, University of Central Florida, Orlando, FL, USA
melkamel@knights.ucf.edu, rabelo@ucf.edu

Morgan C. Wang
Department of Statistics and Data Science, University of Central Florida, Orlando, FL, USA
Chung-Ching.Wang@ucf.edu

Abstract

Artificial intelligence (AI) is continuing to gain traction as technology improves and there is increased data collection. AI can assist society by predicting future phenomena and thus this leads to more efficient processes, increased profits, and lower costs. The evolving world environment is leading to increased pressure for food, energy and water resources and it becomes increasingly important to manage these resources. AI can be utilized in the form of the internet of things (IoT), which are sensors and devices that collect data, to help manage natural resources. In this paper, several machine learning models are employed to predict crop yield. Crop yield is an important factor for many countries and can be an indicator of food security. The ability to predict crop yield is of increased importance as it can aid policymakers in planning future food supply and can aid in the development of a smart irrigation system. A case study of sugarcane in the state of Florida is utilized, along with climate and soil variables. As a follow up to the results of the machine learning models on crop prediction, this paper introduces an optimization model that develops the smart irrigation system. The model chooses the optimal amount of irrigation and fertilizer needed to maximize crop yield while also minimizing costs, water, and energy resources. The machine learning models identify the most important variables needed for crop prediction and can be applied to other crops in other regions of the world.

Keywords
Machine Learning, Crop Prediction, Smart Irrigation System, Internet of Things, Optimization

Biographies
Marwen Elkamel is a graduate student currently pursuing his PhD in the Department of Industrial Engineering & Management Systems at the University of Central Florida. He obtained a Bachelor degree with distinction in Economics with a minor in Management Studies from the University of Waterloo, Ontario, Canada and a Master of Science degree in Management (Business Analytics track) from the University of Central Florida. Before starting his PhD, he worked as a data analyst for WeCare tlc. During his undergraduate studies, he served as a Research Assistant at the Waterloo Institute for Sustainable Energy (WISE). He was involved in two different projects that encompassed the acquisition and summary of data and preparation of computer programs to simulate processes and to make appropriate conclusions. During his PhD studies, he has been preparing machine learning models for electricity consumption with the consideration of socio-economic factors. He was also involved in a project that dealt with power resources scheduling and planning. He is currently focusing on modeling and optimizing the Urban Food-Energy-Water Nexus in order to find more efficient ways to supply water, energy and food and manage natural resources that can aid in sustainable energy development and improved water and food security. He is a member of IEOM, IFORMS, and the Institute of Industrial & Systems Engineering. He has published several journal and conference papers in the areas of modeling, simulation, optimization, and big data analytics.

Dr. Luis Rabelo is a Professor of Industrial Engineering & Management Systems at the University of Central Florida. He was the NASA EPSCoR Agency Project Manager (2009-2011). He received dual degrees in Electrical and Mechanical Engineering from the Technological University of Panama and Master's degrees from the Florida Institute

of Technology in Electrical Engineering (1987) and the University of Missouri-Rolla in Engineering Management (1988). He received a Ph.D. in Engineering Management from the University of Missouri-Rolla in 1990, where he also did Post-Doctoral work in Nuclear Engineering in 1990-1991. In addition, he holds a dual MS degree in Systems Engineering & Management from the Massachusetts Institute of Technology (MIT). He has over 300 publications, three international patents being utilized in the Aerospace Industry, and graduated 35 Master and 24 Doctoral students as advisor and co-advisor. He has consulted with NASA, NSF, ONR, NIST, Lockheed Martin Corporation, Boeing, Tyco, and others. His experience includes Ohio University, BF Goodrich Aerospace, Honeywell Laboratories, the National Institute of Standards and Technology, NASA, and MIT. He has received many awards among them ONE NASA in 2006, the Alumni of the Year of the Technological University of Panama in 2008, Fulbright Scholar in 2008, Two NASA Group Achievement Awards, the Emerald Literati Network Awards for Excellence 2007, the 2004 Arch T. Colwell Merit Award from the Society of Automotive Engineers (SAE), the 23rd Annual Hispanic Engineer National Achievement Awards Corporation (HENAAC) Education Award Winner in STEM in 2011, the Engineer Educator of the year 2011 by the US Engineer's Council, and the 2013 International Joseph McFarland Award from SAE.

Dr. Morgan C. Wang received his Ph.D. in 1991. He is the funding Director of the Data Mining Program and professor of Statistics at the University of Central Florida. He is the elected member of International Statistical Association and member of American Statistical Association and International Chinese Statistical Association. He has published one book (Integrating Results through Meta-Analytic Review Using SAS Software, SAS Institute, 1999), and over 80 papers in referee journals and conference proceedings on topics including interval analysis, meta-analysis, computer security, competing analytics and data mining. He couched student teams to win the 2011 and 2012 SAS Data Mining Shoutout Contest. He won the best conference award in the First Annual Conference on Engineering and Technology Innovation in 2008. He was the first prize-winner in Data Mining Competition of the 11th SIGMOD KDD (the most predigest data mining competition) conference in 2004 and the first prize winner in Data Visualization Contest of SUGI 25 conference in 2000, and was given invited talks on making intelligent decision based on large volume of observational data to more than eighty times for American Statistical Association, SIGKDD (leading conference in data mining), International Conference on Information Technology, SAS Global Forum, Well Fargo Bank, Everbank, Republic Bank, Florida Blue, Disney, Kemper Preferred Auto Insurance, HealthFair, and many companies and universities around the world.