

# A comparative study of machine learning methods for forecasting prevalence of weather-based pests

Too, Edna Chebet

## Abstract:

The aphid (*Aphis gossypii*), have been identified as one of the major pest problem in Kenya (Waiganjo, et al., 2006). They cause damage by sucking plant sap weakening the plants, and by excreting a sticky substance (honeydew), which results in growth of sooty mould affecting photosynthesis. Aphids affect many commercial plants including cereals (maize, wheat, and rice), potatoes, vegetables (cabbages, tomato, okra, and onion). This study concentrates on data for vegetables and specifically Tomato and Okra which are widely cultivated in tropical and subtropical regions. Demand for quality pest-prediction software has undergone rapid growth during the last few years. This has led to increased research in machine learning techniques for exploring datasets which can be used in constructing models for predicting quality attributes. Among the commonly used techniques include Decision Tree (DT) (Pratheepal, Meena, Subramaniam, Venugopalan, & Bheemanna, 2010), Support Vector Machine (SVM) (Kaundal, Kapoor, & Raghaya, 2006), K-means clustering (Al-Hiary, Bani-Ahmad, Reyalat, Braik, & Rahamneh, 2011) and Multilayer Perceptron (MLP) (Worner, Lankin, Samarasinghe, & Teulon, 2002). This study examines and compares Multilayer Perceptron (MLP), SVM, K-means clustering and DT methods. Analysis was done on the two selected vegetables namely Tomato and Okra. Data collected from various sources was used in the study. The data set for studying tomato was obtained and it contained data for the period of 2005-2007 (Chakraborty, 2011) and Okra data set was for the period of 2004-2006 (ANITHA, 2007). The performance of the tools was compared by using DTREG, a prediction modeling software. K-means clustering showed the highest accuracy percentage of classifications of pests with 100% and is a better model than the model predicted using DT, SVM and MLP methods which showed a percentage accuracy of 94%, 60%, and 60% respectively. The finding shows that machine learning methods can be used to construct reliable applications in prediction of aphids based on weather data, with k-means clustering as the most accurate algorithm.