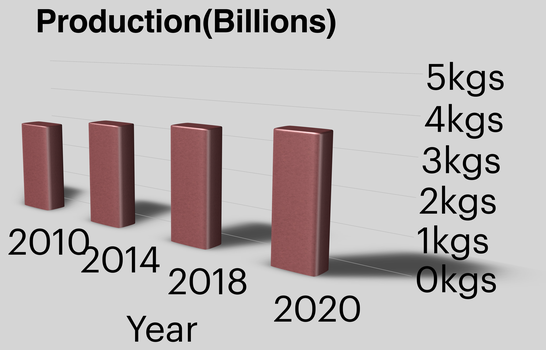
Deriving useful Parameters that may effect the harvest of an apple using Supervised Learning

Ruthvik Raja M.V, Jayasaikiran.A, School of Engineering, University of Guelph

Introduction

- In the last ten years, the annual production of apples in the United States has increased dramatically Thereby, it is essential to identify the key canopy parameters that may influence apple harvesting in order to determine whether an apple is harvested or unharvested.



Literature Cited

1. USDA 2020. National agricultural statistics database.

2. Zhang, X., He, L., Zhang, J., Whiting, M., Karkee, M., Zhang, Q.
New Machine Learning Findings Reported from Washington State University(2020).

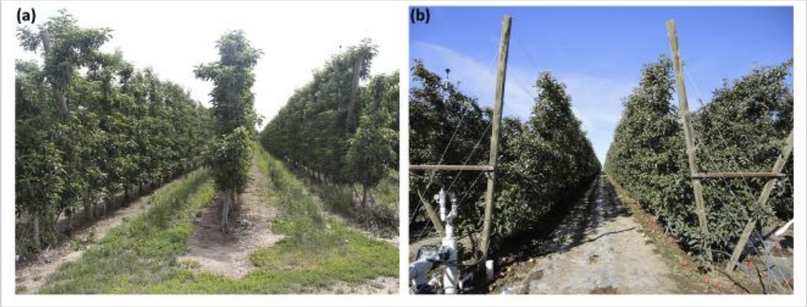
3. Zhang, Q., Pierce, F. (2013). Agricultural automation fundamentals and practices.

Methods

- The apples are grown in a vertically trained "Sci-fresh"(a) and V trellis grown "Envy" trees(b).

- There are 11 parameters that may affect an apple's harvest, including Branch Length, Branch End Diameter, Branch Middle Diameter, Branch Basal Diameter, Fruit Load, Fruit Density, Fruit Location, Fruit Single Mass, Shoot Basal Diameter, Shoot Length, and Shoot Index, but Fruit Load, Branch Base Diameter, and Shoot Length have the greatest impact.

- There are several methods for performing binary classification, but some researchers discovered that by taking into account all 11 canopy parameters, the weighted KNN(K-Nearest Neighbours) algorithm can produce a high accuracy score, and Principal Component Analysis was used to reduce dimensionality.



- Other researchers conducted similar studies and discovered that ANN (Artificial Neural Networks), SVM (Support Vector Machines), and Decision Trees produced high accuracy by considering only the key canopy parameters that actually affect apple harvest.

Results

- The Accuracy scores of the weighted KNN on the testing set for "Sci-fresh" and "Envy" trees by considering all the 11 parameters, ANN and SVM by considering only the key canopy parameters are as follows:

Algorithm	Accuracy Score
KNN("Sci-Fresh")	$(85.9 \pm 0.2\%)$
KNN("Envy")	$(68.5 \pm 0.5\%)$
SVM	99%
ANN	97%

Conclusions

- This poster presented various methods available in Supervised Learning for classifying apples as harvested or unharvested using 11 canopy parameters.

- In the near future, the goal should be to focus only on the key canopy parameters that have an effect on the output variable, such as Fruit Load, Branch Base Diameter, and Shoot Length, in order to achieve a high accuracy score. Instead of Principal Component Analysis, advanced algorithms such as minimum redundancy maximum relevance can be used for feature selection to achieve a higher prediction score.

Acknowledgements

I thank Professor Ahmed Mahmood for suggesting me a problem statement that may help my future Major Research Project and Professor Luiza Antonie for helping me to create a poster on my Literature Survey paper.

Contact Information

For further information please reach us at rmunirra@uoguelph.ca (or) jalluri@uoguelph.ca.