**CKME 136 – Literature review**

Matthew Woods, 500407864**,** matthew.woods@ryerson.ca

**Introduction**

Greenhouses are commonly used to grow seasonal produce and ornamental flowers. As the population in canada increases we should also expect to see an increase in greenhouse production. One common area of research pertaining to greenhouses is the use of statistical models for the use of predicting plant yield. Without the use of statistical modeling, farmers typically rely on experience and intuition to estimate total plant yield, both of which are sometimes unavailable. These models are useful for farmers in two main resects. First and most obviously, they enable farmers to accurately predict yield. Secondly, these models can be broken down and interpreted through their features, such as light intensity, temperature, fertilizer, and day length. This in turn aids farmers in their understanding of which variables cause the greatest change in plant yield.

This project seeks to estimate what will be the total greenhouse sales in future years in Ontario and Canada through the use of statistical modeling and regression techniques. We would also like to explore whether greenhouse features in Ontario (space, energy expenditure, employees) are significantly different than the rest of Canada. Greenhouse industry statistics, which can be found on the Ontario data catalog, contains the dataset which will be used for this project. The data will be analyzed by RStudio using linear regression and two sample t-tests.

**Literature Review**

1. Energy and cost analysis for greenhouse and open-field grape production.

Burhan Ozkan, Cemal Fert, C. Feyza Karadeniz

The purpose of this paper was to study the energy usage of typical greenhouse and open-field grape production. Value was calculated as both yield and total-net worth since both are valuable metrics for farmers. The experiment concluded that with current greenhouse technology open-field planting is more energy efficient because of reduced electricity/lighting demands. Greenhouses however, provide higher profits to farmers due to premium prices for greenhouse produced grapes. Besides premium grape prices, greenhouses call also be run all year round which may provide more growing seasons. This paper was useful for the literature review as it illustrates the relationship between energy efficiency and operation costs for farmers.

1. Yield prediction for tomato greenhouse using EFuNN

Kefaya Qaddoum, E. L. Hines, D. D. Iliescu

This study focuses on creating a tomato yield prediction model in greenhouses through EFuNN (evolving fuzzy neural network). The neural network system in this experiment takes in typical greenhouse parameters (fertilizer usage, lighting) and atypical parameters (employee growing experience) to create the yield prediction model. This is useful for the literature review since the experiment takes a data science approach (EFuNN) to solve a practical agricultural problem (yield prediction).

1. Crop modeling and yield prediction for greenhouse-grown lettuce.

W. C. Lin

The purpose of this experiment was to improve yield prediction through the use of artificial neural networks (ANN) as a model. Factors that were included in the yield prediction model include light intensity, temperature, humidity, and nutrition. These factors were recorded on a weekly basis to improve the accuracy of the data. This study helps the literature review as it displays how the use of artificial neural networks can be used to create predictive models for agriculture. It is also useful as it quantifies which greenhouse features effect yield in the most adverse way.

1. Modeling yield pattern of greenhouse-grown sweet peppers.

W. C. Lin, D. Frey, G. D. Nigh, C. C. Ying

The purpose of this study as to model sweet pepper growth using time-series analysis, neural networks, and regression techniques. Neural network and regression models created in this experiment were similar too other NN yield models, however, this one explored the usage of both positive and negative features to predict sweet pepper yield. The time-series methods also showed that yield size was highly dependent on previous weekly yields. This is useful for the literature review as it shows that time-series analysis cannot be used as our data is only taken annually instead of weekly. It also helps by illustrating that simple regression techniques can be used to create accurate yield prediction models for greenhouse produce.

1. The relationship between energy inputs and crop yield in greenhouse basil production.

R. Pahlayan, M. Omid, A. Akram

This piece of literature examined the relationship between basil yield/production and electric energy consumption in greenhouses. The study also broke down energy consumption into multiple categories including electricity, plastic covering, and fertilizer in order to illustrate which features expend the greatest amount of energy. Data was also collected on an annual basis through face-to-face questionnaires, which is similar to the method used in the Capstone dataset. This paper was helpful with the literature review as it displays the value of breaking down variables into smaller sub-parts in order to better understand feature relevance to agricultural yield.

1. A regression model of dry matter accumulation for solar greenhouse cucumber.

Weitang Song, Xiaojun Qiao

The purpose of this study was to create a simple models which could be used to predict the dry matter yield for cucumbers in solar greenhouses. The researchers approach to solving this problem was through creating simple regression models to predict dry cucumber yield. These models were relatively simplistic since they only contained three features: temperature, day length, and light intensity. Results also showed a high coefficient of determination (0.99) which indicates that the model accounts for a large amount of the variation found in the data. This paper is useful for the literature review as it attempts to simplify regression models without sacrificing its predictive power or accuracy. It is also useful as it illustrates which greenhouse features should be included in other greenhouse yield predictive models.

1. Support vector machines regression and modeling of greenhouse environment.

Dingcheng Wang, Maohua Wang, Xiaojun Qiao

Greenhouses are difficult to model as they have a myriad of factor which could affect plant growth/yield. This paper uses support vector machines to model the greenhouse environment to better help predict produce yield. This can be useful for the literature review as it presents another method of using machine learning to predict the total yield in a greenhouse environment.

**Dataset**

The dataset to be used in this project, greenhouse1.xls “greenhouse industry statistics”, was provided by the Ontario government website https://www.ontario.ca/page/sharing-government-data. Greenhouse industry statistics is an annual record of greenhouse produce sale, maintenance cost, and area( m²) between the years of 1996 to 2016. Almost all attributes found in the dataset will be used by some means except for greenhouse purchase value, direct sales to the public ($000), and other channels ($000), since none of them are relevant to any of the research questions.

**Approach**

Data loading, cleaning, processing, and transformation

PCA analysis for Ontario and Canada greenhouse attributes

T-test of crop type vs. greenhouse location(Ontario/Canada)

Create regression model able to predict crop yield in greenhouses

**Step 1. Data loading, cleaning, processing, and transformation**

The first step of the project involves loading, cleaning, processing, and transforming the data to make it usable to RStudio. Values that are empty with either be removed or imputed at the discretion of the researcher.

**Step 2. Create regression model able to predict crop yield in greenhouses**

Using processed data from step 1 and RStudio, a regression model will be built to enable greenhouse farmers to predict yield.

**Step 3. T-test of crop type vs location (Ontario/Canada)**

RStudio and the cleaned data will be used to t-test whether Ontario and Canada have significant crop type growing preferences.