Analysis of Performance of Field Crops

Rizmeer M.S.M.

Department of Computer Science and Engineering

University of Moratuwa

rizmeer.shafeek.19@cse.mrt.ac.lk

Yapa N.H.N.K

Department of Computer Science and Engineering

University of Moratuwa

nuwanthi.yapa.19@cse.mrt.ac.lk

Yeshani R.B.

Department of Computer Science and Engineering

University of Moratuwa

binuri.yeshani.19@cse.mrt.ac.lk

Abstract

Agriculture is the main form of income generation in Sri Lanka. It would be help if farmers can know what are the crops perform well in Maha season and Yala season before cultivating. So that, they can use their money, time and effort on high productive crops and vegetables. K-mean clustering is used to find which crops and vegetables have similar behavior pattern in production from 2001 to 2012. Different visualization techniques helps to figure out relationships between production of crops and vegetables within ten years, production of crops and vegetables during Maha season and Yala season, which crops and vegetables give high production within small area. Government can use these results to encourage farmers and housewives to grow different crops and vegetables. When considering the behavior of productivity of crops during ten years, most of the crops give similar production during ten years. So that, farmers can predict the performance of crops and vegetables and choose which crops and vegetables should be cultivated. Since farmers are able to choose the crops and vegetables with best performance, they are able to give more focus, time and effort on those crops and vegetables. Directly this study add values to country economy as production goes up, income is also gone up. Also indirectly this study can add some values to personal level as well.

Keywords - Performance Analysis, K-Mean, Correlation

1. Introduction

In agricultural sector, enhancing crop production is very important along with the cost spent. The crop productivity prediction is somewhat difficult task now a days. Farmers were used to take decisions from their past experiences that come from generation to generation. Traditionally farmers were decided to make prediction using specific factors such as weather conditions, geographical needs, water cycles, earth-sun relationships. A vast agricultural knowledge will be needed to expect a profitable harvest, a particular crop needs specific requirements to grow and make a good harvest.

Crop prediction approach is a trending topic recently, and efficient approach for better quality farming for increase revenue. Clustering algorithm is a data mining technique used to extract out useful information for more predictions. In crop production analysis, used to group related data into classes or clusters. Crop yield can be predicted using also linear regression models based on basic agricultural parameters like weather, agricultural parameters, soil moisture, surface temperature, rainfall data, etc.

Overall crop predictions and monitoring have been done a massive impact for whole economic development of the country, specially for developing ones. In Sri lanka “Maha season” falls from September to March of the following year and “Yala season” is effective on May to the end of August.

Our proposed analysis based on for identifying relationships along with the production and with respect to cultivated area, based on the seasonality. And predict and make strategic decisions that will useful for farmers, agricultural authorities and country economic sector.

In the rest of the paper organized as follows section 2 discussed about the data set that used, section 3 about the motivation for the study, section 4 illustrates the proposed methodology for the study, selection 5 demonstrate the experiments and results that obtained from study and selection 6 concludes the paper.

2. Overview of the Data Set Used

Open Data initiative of Government of Sri Lanka is responsible for building different datasets which are freely available to public. So that researchers, policymakers, software developers and general public can use these dataset according their purposes. These datasets are machine readable and well structured. Data and information are national assets of Sri Lankan Government. The data sets are categorized under several domains such as travel, information technology and cyber security, agriculture, economics and etc.

Among several areas, we selected agriculture and livelihood category. There are several datasets are available under this category as well. After going through several datasets, we selected “Performance of Field Crops - Maha Season (2001 - 2012)” dataset. This dataset contains data from 2001 to 2012 under forty crops. Dataset consists of forty rows and twenty columns. The crops can be divided under two main categories such as seasonal crops and vegetables. Basically this dataset is for analysis the performance of field crops throughout ten

years. For that year wise production amount and year wise cultivated area are used. From that we can get an idea about productivity of each crops. This will help to identify the products which has high productivity within less area.

Since this dataset is small and only contain the information about Maha season, we used similar dataset as "Performance of Field Crops-Yala Season (2001 - 2012)". This dataset is more similar to above mentioned dataset. Only difference is that it contains the production information related to Yala season. So that we can analyze the production between two seasons.

The both datasets were published by Department of Census and Statistics in 2017. And both datasets were modified in 2018.

3. Motivation

In Sri Lanka, agriculture is main form of income generation. Crop prediction is emerging topic in current era, during the rising reasons. Cultivators are interested curious about how much yield can he gain from cultivation.

Climate is changed time to time throughout the year. Crops had a popular season, that maximize the production. Rice, Crops and other products are mainly cultivated during Maha and Yala seasons. Among these two seasons, farmers tend to grow more crops and others in Maha season. Analysis on performance of crops during seasons will helps to farmers for getting an idea about production at the end of the season. So that they can decide what kind of crops or vegetables should be cultivated to get more production. Farmers will able to afford more time, money and effort on the crops which will give high productivity without wasting time,money and effort on less productive products. Sri Lankan government can also give instructions to farmers about the crops and vegetables by mentioning these products will give high production.

Most of the Sri Lankan people don't have large lands to cultivate. So only less amount of people tends to cultivate within their premises. Identification of crops and vegetables which give high productivity from small land or area will helps to encourage more people on cultivation. Mainly housewives can use these analysis and check about the crops and vegetable which gives high productivity. So that they can grow such crops and vegetables and include fresh vegetables and crops in their meals without pesticides.

This study can be used for directly in economic growth. Indirectly, this can be used to reduce the personal health issues as this helps to eat fresh vegetables and crops.

4. Methodology

In this paper, a study on clustering approach for agriculture to analyse the data set using k-means, linear regression algorithms using python programming. The overall picture of the research scenario will be as following diagram.

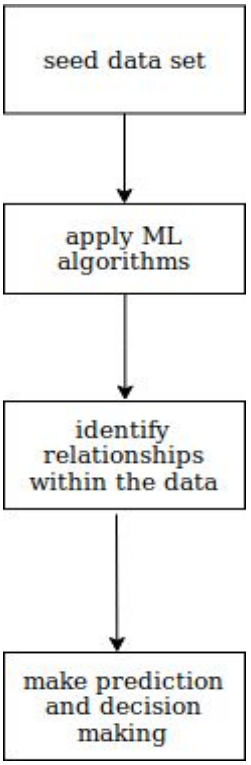

Figure 4.1 overview of the process of the analysis

4.1. Seed Data Set

Initial Data set was about the crops production and cultivated area that measured with respect to metric tons and hekteyar along with the year. Agricultural economics measures farms productivity, production per unit of land that was used for cultivation.

4.2 Apply Machine Learning Algorithm

k-means is an unsupervised, non deterministic and a gathering technique used for clustering purpose. Clustering used to identify which crops are interrelated and behaves similarly that implies which crops are most suited for our country for make more profit. And apply linear regression, correlation to identify which factors that are highly impact for the crops growth.

4.3 Descriptive Analytics,

Descriptive statistics used for reporting and to identify patterns of crop production, it explicitly shows that what has happened in the past.

4.4 Predictive Analytics,

Make prediction about unknown future events that uses for many techniques statistical modeling, machine

learning, modeling to analyze current data to know future in present.

5. Results and Discussion

The dataset is evaluated under different parameters such as production of field crops during ten years, productivity of field crops with compare to area cultivated, production of field crops during Maha season and Yala season and correlation between year wise total production of field crops.

First we try to analyze the how is production of different crops through ten years in Maha season. This will helps to identify the crops with better performance in Maha season. Sri Lankan government and farmers can target on those crops in Maha season. Sri Lankan government should give instructions what kind of crops should be cultivated in Maha season instead of cultivating with low production. So that farmers can gain more without wasting time and money on less productive crops.

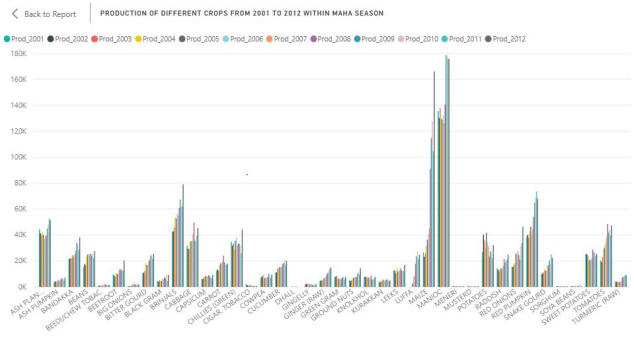

Figure 5.1 Performance of Production of Field Crops in Maha Season from 2001 to 2012

According to above chart, manioc has high productivity than other products. If farmers can give more effort on cultivating manioc than other crops, they will be able to earn more. As per the chart, we can consider as time and effort which is used to cultivate chew tobacco, cigar tobacco, dhal, meneri, mustard, sorghum and soybeans is less valuable.

Most of the Sri Lankan people own very limited area to live, cultivate and all other purposes. Because of this reason, most of people do not tend to cultivate. As a solution to this context, we analyze are there any crops which gives high productivity using less area. Below chart illustrates the above context.

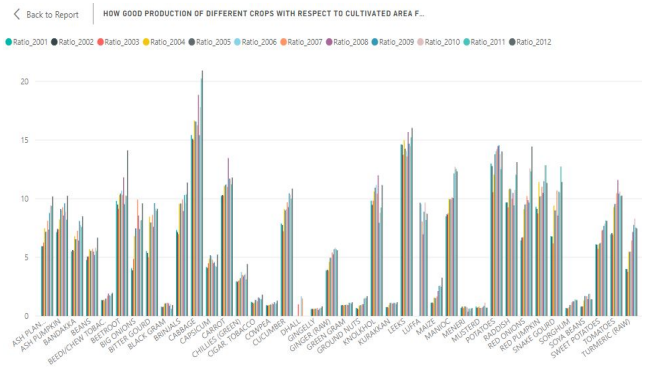

Figure 5.2 Production of crops with respect to Area cultivated in Maha Season from 2001 to 2012

According to the above chart, it can be seen that cabbage have high ratio between production and area cultivated. That means farmers can give high productivity of cabbage within less area. This analysis can be used by government to encourage housewives to grow cabbage, leeks, beetroot and red onion within their house premises. So that they will be able to use vegetables and crops without pesticides.

So far we have discussed only considering Maha season. Below chart will illustrates the comparison between production of field crops during Maha season and Yala season within 2001 year. Production of Maha season is higher than production of Yala season. So farmers can more focus on Maha season than Yala season

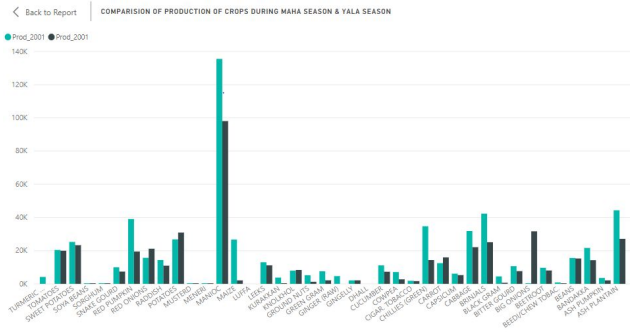

Figure 5.3 Comparison between production in Maha season and Yala season

Our next approach is to get the correlation between year wise production in Maha season. Correlation is a statistical technique that can show whether and how strongly pairs of variables are related. It can be seen that, there is a great correlation between ten years. That means production of field crops are almost similar during Maha season.

Three clusters are identified using K-means clustering algorithm as follow. It illustrates how forty field crops are clustered into three clusters within 2001. Values are the ratio between production amount and the cultivated area. Among three clusters, first cluster has nine crops or

vegetables, second cluster has sixteen crops or vegetables and third cluster has fifteen crops or vegetables.

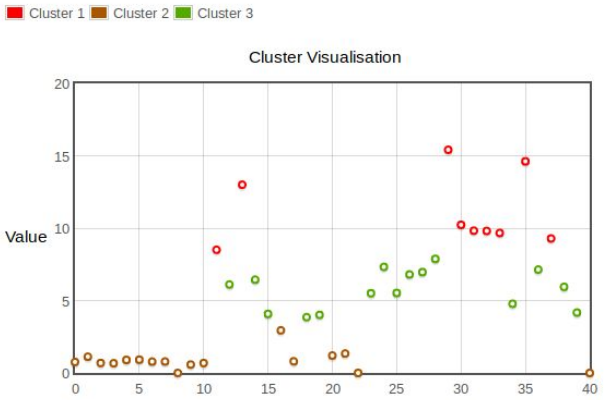

Figure 5.4 K-Means Cluster Visualization

Below table illustrates crops and vegetables belong to three clusters.

Cluster 1	Cluster 2	Cluster 3
Manioc Potatoes Red Onion Mustard Ginger (Raw) Tobacco Luffa Bandakka Brinjals Bitter Gourd Tomatoes Raddish Leeks Red Pumpkin Ash Pumpkin	Kurakkan Maize Sorghum Meneri Green Gram Cowpea Soya Beans Black Gram Dhall Gingelly Big Onions Chillies (Green) Turmeric (Raw) Cigar. Tobacco Beedi/Chew Capsium	Ground Nuts Sweet Potatoes Cucumber Cabbage Carrot Knolkhol Beet root Beans Ash Pumpkin

Table 5.1 List of Crops and Vegetables under Clusters

6. Conclusion and Future Works

In our data set included crops production along with respect to year, if data will be spread out within the whole year can be extract out the seasonality prediction/time series prediction, how will be the production deviation goes on in whole year.

We were able to identified the most productive crop or vegetable among ten years. So that we can encourage farmers to grow those most productive crop or vegetables. When farmers grow the most productive crops or vegetables, they can earn money and help to improve the economic of the country. By identifying crops or vegetables which give more production within small area, we can encourage housewives to cultivate those products within their premises. So that they will be able to include fresh vegetables and crops into their meals.

As further work of this study, we can join annual rainfall rate how will be affect for the crops production. That will be gibe good analysis which is needful to farmers.

Acknowledgement

The authors feel a deep sense of gratitude to Dr. Uthayasanker Thayasivam,Senior Lecturer of Department of Computer Science and Engineering, and Dr.A.Shehan Perera, Head of the Department of Computer Science and Engineering, for their guidance, inspiration and motivation throughout this work.

References

[1] Utkarsha P. Narkhede, K.P.Adhiya, "Evaluation of Modified K-Means Clustering Algorithm in Crop Prediction," in *International Journal of Advanced Computer Research*, 2014.

[2] Shreya S.Bhanose, Kalyani A. Bogawar , Aarti G. Dhotre, Bhagyashree R. Gaidhani,, " Crop and Yield Prediction Model," *International Journal of Advance Scientific Research and Engineering Trends,* 2016.

[3] Dr Madhavi Gudavalli, Vidyasree P, S Viswanadha Raju, "Clustering Analysis for Appropriate Crop Prediction using Hierarchical,Fuzzy C-Means, K-Means and Model based Techniques," *International Journal of Advance Engineering and Research Development,* 2007.