# Prediction of Tea Production in Kenya Using Clustering and Association Rule Mining Techniques

## Nzuva Mutie Silas* and Lawrence Nderu

Department of Computing and Information Technology, College of Pure and Applied Sciences Jomo Kenyatta, University of Agriculture and Technology, Kenya

*Corresponding author: Nzuva Mutie Silas, Department of Computing and Information Technology, College of Pure and Applied Sciences Jomo Kenyatta, University of Agriculture and Technology, Kenya, E-mail: nzuvah@gmail.com

## Abstract

As at present, the agricultural sector is the backbone of the Kenyan economy. Though there has been a significant focus on other emerging industries, the agriculture sector remains to be a crucial player in the Kenyan economy, and which vastly contributes in the provision of job opportunities for millions of Kenyan citizens as well as strengthening the Gross Domestic Product (GDP). Therefore, efforts towards strengthening this sector are highly and warranted. Mining the past agricultural data to establish any new knowledge is, hence of great essence. Knowledge discovery is a crucial component of the modern day decision making. In the agricultural sector, the knowledge gained from the past data can be used for various beneficial purposes, including planning, budgeting a forecasting the possible future production trends. This paper attempts to predict tea production in Kenya through step-wise use of the clustering and association rule data mining techniques. A conclusion is presented, based on the presented arguments.

**Keywords:** Data mining; Association rule; Clustering

## Introduction

**Background**: The 21st century has seen a vast revolution, characterized by tremendous technological advancement. Owing to this, data mining has emerged to be an area of interest that has attracted a significant number of researches and experimentations in an effort to enhance efficiency and productivity in various disciplines. Data mining can be defined as a process that entails establishing previously unknown, non-trivial and hidden information in a larger dataset [1]. The established relation-ships, associations and patterns within the data can aid in providing useful information [2]. In line with this, Geetha argues that data mining tasks can be categorized in to two; predictive data mining and descriptive data mining [3]. The descriptive category entails mining the general attributes of the data to gain a broader understanding of the different variables, while the prediction category entails the use of one or more attributes of the data or files in a given database to predict future or unknown values of same or related variables [3].

Yethiraj review on the application of data mining concepts in the agricultural disciplines establishes that there are different techniques and algorithms that can be employed and specifically, points out the techniques such as K-Nearest Neighbour (KNN), Iterative Dichotomiser 3 (ID3), Artificial Neural Network (ANN) do support vector machines in the agricultural domain [4]. On the same, Raoranne and Kulkarni critically discuss how data mining can aid in boosting agricultural production, through crop yield estimation among others applications. This clearly implies that by mining agriculture-based data, it is possible to love complex problem that face the agricultural industry. As such, clustering and association rule data mining techniques remain an important approach to knowledge discovery. Khan and Singh defines the association rule as an important set of method that is used to establish patterns or regularities and clustering as categorization of the dataset records in to clusters, based on the specific attributes [5].

**Problem statement**: Being a third world country, it is clear that Kenya highly depends on agricultural sector, and enhancing food production in the country remains a top priority. The main aim of this paper is to conduct a comprehensive data mining on the Kenyan tea production using clustering and association rule data mining techniques. The data-set used is from the Kenya Open Data on National Monthly Production, Consumption and Export of Tea in 2003-2015. Specifically, the study is aimed at establishing any existing association and patterns in production during the aforementioned period and uses the established patterns to predict future tea production. The main focus is to outset any relationship in tea production of different months of the year, from 2003 to 2015. In order to enhance tea production, plan for the future production and increase profitability of the ventures, the tea farmers need to understand the trends in the production, consumption and export. At present as at present, technology has become the backbone of business operations and all the industries have become reliant on the same.

While the agricultural sector in kenya is gradually adopting technology in management of different operations, majority of the farmers are yet to adjust, and the ministry has however lagged behind in providing the necessary information to the farmers to enhance agricultural production. Therefore, if the farmers and the government at large continue investing in tea production in particular, and agricultural industry in general without clearly understanding how production can be enhanced in the future by using knowledge from the past data, not only would this amount to wastage of vital resources, but also would jeopardize the main aim of having a functional agricultural sector. This research paper is made to bridge this knowledge gap, by utilizing the clustering and association rules to mine data and establish patterns and other regularities that can aid in the prediction of the future trends and productivity [6-18].

## Literature Review

In their study on crop estimation and production forecasting, Benedettia et al. clearly explain the process that can be used for estimating the intended variables, such as crop yield and land use, and takes into consideration various constructs such as the research sample standard deviations and descriptive. Nevertheless, their described procedure is seemingly complex, making it relatively necessary for users of such a procedure to have a statistician and an appropriate computational system. On the same note, Gayam mines the Indian Soybean and Sugarcane production data to establish the crop yield patterns and distribution. The author used qualitative analysis techniques and the Lillie fore approach to test the null hypothesis, Crop yield are normally distributed [10]. The results of the study indicated that there is no normal distribution in the Indian Soybean and Sugarcane yields. In an effort to enhance data mining in the agricultural disciple, Yethiraj carried out an investigation on how to best apply various techniques in mining agricultural-related data.

The researcher presents various data mining techniques and their application in the agriculture and allied disciplines, such as the Support Vector Machines, Artificial Neural Net-works, the K-Nearest Neighbour technique, the K-means, Iterative Dichotomiser 3 algorithms and Association Rule Mining technique. The researcher denotes that in foresting prediction of agricultural crops and animal production is relative a new area of interest, but however, there is no one best technique, and the appropriateness of the method chose depends on the data set being mined, the expected results and the researchers knowledge about the technique [4].

Yethiraj systematic review on data mining techniques in the agricultural domain established that there are numerous algorithms that can be employed to recognize patterns in the data and aid in future prediction in production [4]. This is as well affirmed by Barghavi and Jyothi who claims that data mining techniques can be employed in a given agricultural dataset to gain information, though this highly depends on the amount of data used [13]. This clearly implies that accuracy of the gained information can be enhanced by increasing the size of the data-set. Such an aspect may enhance the valid patterns verification as compared to the conventional statistical analysis.

The data to be used in a given study may be obtained from numerous sources, and hence the need to employ the most efficient and effective data mining technique [14]. This implies that there is the need to carry out various processes for the chosen algorithm to work perfectly. Vibha and Yashovardhan assert that there is the need for prepossessing such as data cleansing and integration before. During this phase, the irrelevant data and noise data are eradicated from the given dataset [14]. Data transformation entails the transformation of the selected data into a form that can appropriately analyze by a given data mining technique and establish potentially useful patterns. On a different study to establish corn performance, Lansigan simulated corn performance by employing simmeto, a weather generator model for the generation of daily weather data sequences and changing climate scenarios with respect to the foreseeable seasonal changes in the climate conditions

The researcher then used multiple regression analysis in predicting the yield of the corm, through the use of climatic variables like solar radiation, wind speed, rainfall, humidity, and temperature [16]. The study findings indicated that the crop yields were significantly impacted by climate variability.

According to Raoranne and Kulkarni discussion on data mining in the agricultural discipline, the authors explain that the latter can greatly aid in linking the knowledge gained from the mined data to agricultural yields estimation [8]. This is affirmed by Vamanan and Ramar who asserts that classification approach in data mining can be applied to soil and crop datasets to establish any meaningful association between variables in the dataset [9]. Raoranne and Kulkarni applied different mining techniques on the identified variables to outset the existence of any meaningful relationships [8]. It was established that there is the need to develop more efficient techniques for mining and analyzing the given data in an effort to solve sophisticated agriculture-based problems. On a different note, Genetic Algorithm has been found by various researchers to be an effective data mining technique that can aid in pattern recognition in a given dataset. Nevertheless, Hassani and Treijis express concern over the premature convergence of the technique, an aspect that may thwart population diversity and inhibit the entire search space exploration [19]. Having identified the problem, Hassani and Treijis proceeds to suggest correctly set all the given parameters and Genetic Algorithm tweaking to match a particular problem [19]. This is exemplified by Ying et al. whose study unveiled promising results, based on negative section; a tweak of the genetic algorithm [20].

Data mining techniques have widely been applied in different domains. Specifically, association mining has proved to be very useful specifically in business field in the discovery of purchase patterns, sales patterns and the association between different commodities, and objects [6]. Owing to this, it is clear that association mining can greatly aid in discovering new knowledge for the agricultural data set and aid in boosting production and profitability of agriculture based ventures [1]. Geetha asserts that data collected form agricultural surveys regarding crop production, geographical conditions, soil and cultivation can be mined to establish any regularity in that data and which can be used to predict future aspects [3].

In his study on spatial data mining in the agricultural discipline, Rajesh surveys the different clustering analysis techniques that are aimed at establishing rules and patterns in the given dataset [17]. According to Rajesh, association rule mining in large datasets that is aimed at outsetting spatial relationships among the objects could be cumbersome, thereby calling for a technique called progressive refinement. The researcher achieved this through the use of the K-Means algorithm [17]. Bhatia and Anu Gupta used an agricultural data warehouse to mine quantitative association rules ion the dataset. The researchers compared different association rule techniques including FP-tree growth algorithm, Dynamic Item Set, Pincer-Search Algorithm, Partition Algorithm and Apriori Algorithm. The authors found the FP-Tree growth algorithm to generate the best results, as the Katter bears the least time complexity, making it more optimal [15]. As at present, the agriculture based datasets are growing by day and the need to explore the data and in to knowledge and convert it to knowledge is ever growing. Essentially, in order to generate knowledge from massive database, it is crucial to select the most appropriate technique when dealing with agricultural database in an effort to discover new knowledge and when dealing with different attributes [21].

In summation, the have been found to be highly useful especially when dealing with large agricultural production data-sets in the agricultural domains. This implies that the technique can be applied to a data set to outset unknown patterns relative to the association of the variables over a given time and under different conditions. As informed by studies from prior researches, there is no one best and must use technique, and therefore, the choice of the approach used depends on the applicability to data-set to be mined, its convenience to the researcher and the expected results of the study

## Methodology

**Clustering**: This entails the categorization of the dataset records in to clusters, based on the specific attributes. Clustering therefore breaks down a dataset into subsets, whereby the varied elements are assigned to varied groups while the similar instances are grouped together. According to Rokach, clustering is usually done with the aim of establishing different categories in a given dataset, thereby giving the researcher more insights on the data [12]. In this study, a K-Means clustering Algorithm was used. The latter can be defined as an iterative algorithm whereby the elements are partitioned into K clusters. An element therefore belongs to the cluster with the most proximal mean [3].

D=t1, t2, t3,...,tn//Set of elements k//

Number of desired clusters Output: K//set of clusters

**Association rule mining**: This research relied on the use of the association rule in data mining to establish all the co-occurrence relationships, which can, be perceived as associations. According to Khan and Singh, market based data analysis is the most common application of the association rule, but however, this technique can be applied in other data-sets whereby the researcher wishes to outset new patterns [7]. Essentially, the association mining rule exist in the form X implies Y, where X and Y are collection of variables in a dataset and the intersection of Y and X is null [1].

**Association rule mining definition of the steps**

D = t1, t2,…,tn is a dataset containing tea production in Kenya from 2003 to 2005.

Each record consists of I, where i1, i2, in=I is a set of all items.

Association rule mining entails the implication of the form AB, where A and B are item sets, AŁI, BŁI, AB=.

Each association rule established encompasses support and confidence aspects.

The support denotes the occurrence rate of an item set in D, and the confidence denotes the proportion of data items containing B in all items containing A in D.

Sup(k)=Count(k)/Count(DBT)  Sup(AB)=Sup (Ał B)  Conf (AB)=Sup(Ał B)/Sup(A)

When the support and confidence are greater than or equal to the pre-defined threshold Supmin and Confmin, the association rule is considered to be a valid rule

**Apriori algorithm**: There are different algorithms that have been developed to enhance association rule data mining technique. Any given algorithm should offer high computational efficiency at the lowest memory requires possible [5]. The algorithm used entailed two main steps:

Generation of the frequent item sets: A frequent item set can be perceived as an item set whose transaction support is seemingly above the minimum support.

Generate the confidence association rules from the identified tensest: This is a rule whose confidence is above the minimum confidence. As advised by Patel and Patel, apriori algorithm utilized will highly depend on the downward closure or apriori to generate the frequent item sets in the given tea dataset (Figures 1-9).

## Analysis

**Clustering**: The data-set comprised of 156 tea production records obtained from Kenya Open Data on national production and consumption of tea from 2003 to 2015. The main aim of clustering was to group together the months with seemingly similar productions. This was achieved through the use of the Statistical Package for Social Sciences (SPSS) K-Means component. The data was grouped into four initial cluster centers and iterated seven times to obtain the final cluster centers. A Convergence was achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center was .000. The total number of iterations was 7.

**Initial Cluster Centres**

| | Cluster | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Millions in KGS | 34.4 | 23.2 | 45.4 | 11.7 |

**Final Cluster Centres**

| | Cluster | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Millions in KGS | 32.8 | 25.9 | 40.3 | 18.7 |

**Figure 1:** The minimum distance between initial centers was 11.000 is presented in the figure.

Based on the K-Means clustering algorithm, the data-set was broken down into subsets, whereby the varied elements were assigned to varied groups while the similar instances were grouped together. This is in accordance to Rokach that clustering is usually done with the aim of establishing different categories in a given dataset, thereby giving the researcher more insights on the data [12]. The first cluster contained 53 records (months), which had production values closer to 32.8 Million Kg, the second cluster comprised of 49 records (months). These had their total close to 25.9 Million Kg. The third cluster comprised of records with values closer to 40.3 Million Kg and was 31 records (months).

**Number of Cases in each Cluster**

| | | |
|---|---|---|
| Cluster | 1 | 53.000 |
| | 2 | 49.000 |
| | 3 | 31.000 |
| | 4 | 23.000 |
| Valid | | 156.000 |
| Missing | | .000 |

**Figure 2:** The final cluster comprised of 26 records (months) and had values closer to 18.7 Million Kg is presented.

A cluster graph was developed to show the different clusters. However, it was unveiled that no definite clusters could be clearly established as there was the absolute distribution of the points on the graph, but following a specific pattern. From the graph below, it is notable that while there is no clear distinction or boundary of the clusters as established in the cluster figures, there is a definite rise in trend, with the majority of the points lying below the 30.0 mark on the Y-axis, but increasing across the graph. This was attained by developing a quantric graph of the different points through-out the graph.

**Deduction**: An extrapolation of the quadratic graph from the scatter positions of individual records indicates the possibility of increased production. While the total amount cannot be precisely defined, it is observable that holding other factors constant, future production is likely to be above average, based on the cluster patterns of the past 13 years.
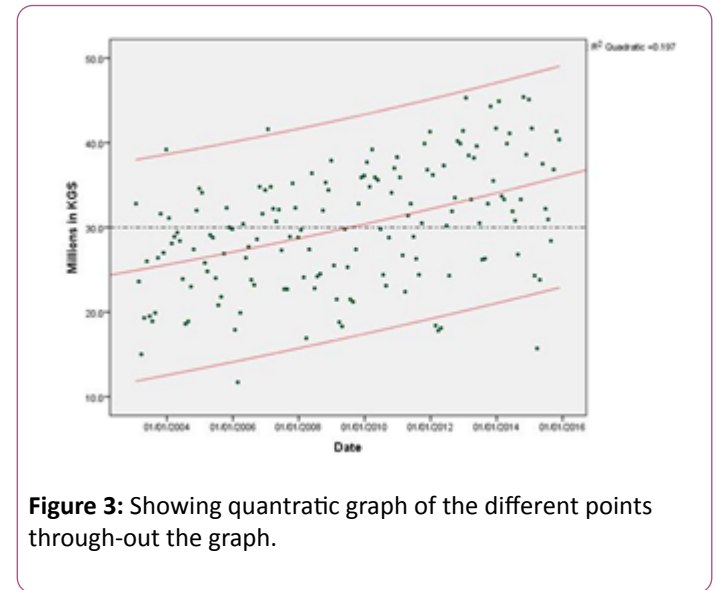


**Figure 3:** Showing quantratic graph of the different points through-out the graph.

**Association rule**: Data Summarization: As defined earlier on, association mining is specifically made to establish an association between variables, based on their rates of occurrence. As explained, the data-set used in the study comprised of 156 records (Months). However, in order to establish the rules, the data was had to be summarized and logical implication introduced. In order to summarize the data, four main variables were established:

Extremely high: This represented the months with extremely high production.

High: This variable represented the months with high (less than extremely high) but more than average.

Average: This represented the months with average production.

Below average: This represented the months with production below average but above extremely low production.

Extremely low: This represented production that is less than below average production

1. Production>45⇒ Extremely High
2. 35< Production <=45⇒ High
3. 25< Production <=35⇒Average
4. 15< Production <=25⇒Below Average
5. Production < 15⇒low

**Figure 4:** Criteria and Implications: The criteria entailed classification of the data and establishment of the classes with boundaries.

**Weighting of the monthly production and logical implications:** Each of the four variables was transformed by assigning those weights in order to make it possible to analyse the variables, based on the monthly production.

**Implications:** With regard to the production recorded across the 156 months in the 13 years, each of the months was assigned a value (weight), depending on whether the production was Extremely High, High, Average, Below Average, Low, or Extremely Low.

1. Production>45⇒ Extremely High⇒5
2. 35< Production <=45⇒ High⇒4
3. 25< Production <=35⇒Average⇒3
4. 15< Production <=25⇒Below Average⇒2
5. Production < 15⇒low⇒1

**Figure 5:** Showing the results of the summarization and variable transformation.

|  | J | F | M | A | M | J | J | A | S | O | N | D |
|------|---|---|---|---|---|---|---|---|---|---|---|---|
| 2003 | 4 | 5 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 3 | 4 |
| 2004 | 4 | 3 | 3 | 2 | 3 | 4 | 2 | 2 | 3 | 2 | 3 | 3 |
| 2005 | 4 | 3 | 3 | 3 | 3 | 4 | 2 | 2 | 4 | 3 | 3 | 3 |
| 2006 | 3 | 3 | 3 | 4 | 3 | 4 | 2 | 3 | 3 | 3 | 3 | 3 |
| 2007 | 3 | 3 | 4 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 2 |
| 2008 | 3 | 4 | 4 | 2 | 2 | 3 | 4 | 4 | 3 | 3 | 2 | 2 |
| 2009 | 4 | 4 | 3 | 2 | 4 | 4 | 3 | 4 | 3 | 2 | 2 | 2 |
| 2010 | 2 | 3 | 4 | 4 | 4 | 4 | 4 | 3 | 2 | 2 | 2 | 3 |
| 2011 | 2 | 3 | 5 | 4 | 4 | 3 | 3 | 2 | 2 | 2 | 3 | 2 |
| 2012 | 2 | 4 | 4 | 4 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 2 |
| 2013 | 4 | 4 | 4 | 4 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 2 |
| 2014 | 5 | 4 | 4 | 3 | 2 | 2 | 2 | 3 | 3 | 2 | 3 | 3 |
| 2015 | 4 | 4 | 3 | 2 | 2 | 3 | 4 | 3 | 2 | 2 | 3 | 3 |

**Figure 6:** Showing clearly that some patterns in production occurrences have emerged.

However, to visualize this more clearly, the radar and Y, X scatter graphs. The graphs clearly show the fluctuation of the four variables over the 13 years and across the 12 months in every year.
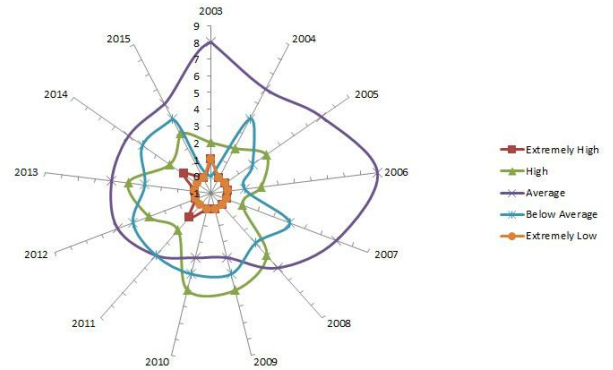


**Figure 7:** Showing the mining occurrences.

Mining the Occurrences: This step entailed establishing the number of occurrences of each of the variable. It is worth noting the variables which include Extremely High, High, Average, Below Average, Low, or Extremely Low. The figure 8 below summarizes the occurrences recorded, based on the variables and for each of the 13 years irrespective of the month.

| Year | Extremely High | High | Average | Below Average | Extremely Low |
|------|----------------|------|---------|---------------|---------------|
| 2003 | 1 | 2 | 8 | 0 | 1 |
| 2004 | 0 | 2 | 6 | 4 | 0 |
| 2005 | 0 | 3 | 7 | 2 | 0 |
| 2006 | 0 | 2 | 9 | 1 | 0 |
| 2007 | 0 | 1 | 7 | 4 | 0 |
| 2008 | 0 | 4 | 5 | 3 | 0 |
| 2009 | 0 | 5 | 3 | 4 | 0 |
| 2010 | 0 | 5 | 3 | 4 | 0 |
| 2011 | 1 | 2 | 4 | 4 | 0 |
| 2012 | 0 | 3 | 5 | 4 | 0 |
| 2013 | 0 | 4 | 5 | 3 | 0 |
| 2014 | 1 | 2 | 5 | 4 | 0 |
| 2015 | 0 | 3 | 5 | 4 | 0 |

**Figure 8:** Summarizes the occurrences recorded, based on the variables and for each of the 13 years irrespective of the month.

**Derived Set of Occurrences:**

- 2003 (Extremely High, 2 High, 8 Average, 1 extremely low)
- 2004 (2 High, 6 Average, 4 Below Average)
- 2005 (3 High, 7 Average, 2 Below Average)
- 2006 (2 High, 9 Average, 1 Below Average)
- 2007 (1 High, 7 Average, 4 below Average)
- 2008 (4 High, 5 Average, 3 below Average)
- 2009 (5 High, 3 Average, 4 Below Average)
- 2010 (5 High, 3 Average, 4 Below Average)
- 2011 (1 Extremely High, 2 High, 4 Average, 4 Below Average)
- 2012 (3 High, 5 Average, 4 Below Average)
- 2013 (4 High, 5 Average, 3 Below Average)
- 2014 (1 Extremely High, 2 High, 5 Average, 4 Below Average)
- 2015 (3 High, 5 average, 4 Below Average)

The average occurrence of each of the weights is then used as the benchmark to establish any association for any occurrences (production per month) that surpasses the averages. The averages have been calculated by adding all variable occurrences across the 13 years then dividing by the number of years to obtain the average occurrence (number of production months) in every year for the four variables:

- Average for Extremely high=0.23
- Average of High=2.9 months
- Average of Average=5.38 months
- Average of Below Average=3.15 months
- Average of Extremely Low= 0.0769 months

Support and Confidence For:

- Extremely High: No single variable that meets the average threshold
- High: Support: 7/13=0.538; Confidence: 13/13=1
- Average: Support: 5/13=0.384; Confidence=13/13=1
- Below average Support: 8/13=0.615; Confidence: 12/13=0.923
- Extremely Low: No single variable that meets the aver-age threshold.



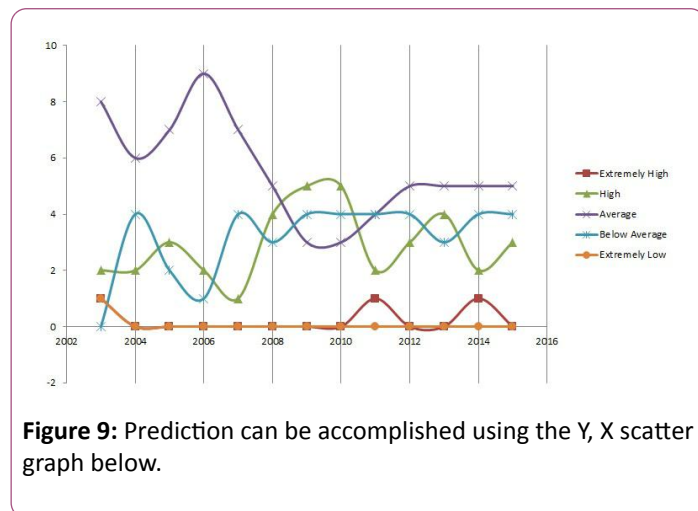**Figure 9:** Prediction can be accomplished using the Y, X scatter graph below.

## Conclusion

Data mining in the agricultural discipline can be perceived to be a novel undertaking; a practice that is targeted at improving the general welfare of the people. Owing to this, yield prediction remains a crucial element that cannot be understated. Essentially, every farmer is highly interested with the level of yields to expected in a given month or season. Earlier on, prediction of crop yield was performed through analysis of the past experience by the farmer. However, in the modern era, data mining procedures can be employed to perform predictions on possible future yields.

Being a developing country, the vision to meet the food demand and alleviate poverty remains a top priority, not only by the Kenyan government but also by the general public. Such an aspect necessitates the use of computerised emission tools and models to streamline production efficiency by the farmers. Such

an aspect has made it necessary to combine the agriculture with advanced technologies to boost crop yields, thereby making predictive models more vital elements.

From the mined data, over the last couple of years, Average tea production has emerged to be the highest trend in the majority of the months. Again, the Extremely Low production and Low production are rarely realized. On the same, there has been a relatively fluctuating Below Average production in different months across the 13 years surveyed, though the span of the yields of the months are lower than that of the Average production but higher than the span of the High production months across the thirteen years.

The High production is highly fluctuating but at decreasing rate. From the association rules established above and based on the curve and figures presented, it is definitely clear that future production will mainly be described by three main variables; High, Average, and Below Average. The Average production is expected to span for the longest period, closely followed by Below Average production and lastly, the High production. Based on the trends identified in figure, the Average production is highly likely to span around February, May, June, September, October and November.

The High production is likely to span for January and December, while the Below Average production March, April, July and August.

## References

1. Patel H, Patel D (2014) A Brief survey of Data Mining Techniques Applied to Agricultural Data. Inter J Computer App 95: 6-8.

2. Ganesh H, Cindrella P, Christy J (2014) A Review On Classification Techniques Over Agricultural Data. Inter J Computer Sci Mobile Comp 4: 491-495.

3. Geetha G (2015) Implementation of Association Rule Mining for different soil types in Agriculture. Inter J Adv Res Computer Comm Engineering 4: 520-522.

4. Yethiraj NG (2012) Applying Data Mining Techniques in the field of agriculture and allied sciences. Inter J Business Intelligents 1: 2.

5. Fathima N, Geetha R (2014) Agriculture Crop Pattern Using Data Mining Techniques. Inter J Adv Res Computer Sci Software Engineering 4: 1-6.

6. Kaur M, Gulati H, Kundra H (2014) Data Mining in Agriculture on Crop Price Prediction, Techniques and Applications. Inter J Computer App 99: 1-3.

7. Khan F, Singh D (2014) Association Rule Mining. Inter J Sci Res Pub 4: 1-4.

8. Raoranne AA, Kulkarni RV (2012) Data Mining: An effective tool for estimation in the agricultural sector. Inter J Emerg Trends and Technol Computer Sci 1: 75-79.

9. Vamanan R, Ramar K, (2011) Classification of agricultural land soils a data mining approach. Inter J Computer Sci Engineering 3: 379-384.

10. Gayam N (2006) Risk in agriculture: a study of crop yield distributions and crop insurance, Doctoral dissertation, Massachusetts Institute of Technology.

11. Benedettia R, Remo C, Federica P (2006) Generalized Software Tools for Crop Area Estimates and Yield Forecast. In Baruth B Royer A Genovese G 10: 7-10.

12. Rokach L, Maimon O (2009) Clustering Methods. Chap-15.

13. Barghavi P and Jyothi S (2009) Applying nave bayesdata mining technique for classification of agricultural land soils, Inter J Computer Sci Network Security 9: 117-122.

14. Vibha M, Yashovardhan K (2012)Survey on Data Mining. Inter J Emerg Technol Adv Engineering 2: 275-278.

15. Bhatia J, Gupta A (2014) Mining of Quantitative Association Rules in Agricultural Data Warehouse: A Road Map. Inter J Info Sci Intelligent System 3: 187-198.

16. Lansigan F (2010) Analysis of climatic risk and coping strategies in two major corn-growing areas in the Philippines.

17. Rajesh D (2011) Inter J Computer App 15.

18. Ramesh M, Vijay R (2014) A survey on Data Mining Techniques for Crop Yield Prediction. Inter J Adv Res Computer Sci Management Studies 2: 1-6.

19. Hassani A, Treijis J (2009) Overview of standard and parallel genetic algorithms, in Proc. IDT Workshop on Interesting Results in Computer Science and Engineering, Mlardalen University, Sweden.

20. Ying LB, Na LN, Hua GJ (2006) A new genetic algorithm based on negative selection. International Conference on Machine Learning and Cybernetics 4297-4299.

21. Thakkar R, Kayasth M, Desai H (2014) Rule Based and Association Rule Mining On Agriculture Dataset. Inter J Innovative Res Computer Communication Eng 2: 1-2.