

MULTI-LEVEL ASSOCIATION RULE MINING FOR THE DISCOVERY OF STRONG UNDERREPRESENTED PATTERNS THE CASE STUDY OF SMALL DAIRY FARMS IN TANZANIA

Glory C. Malamsha , Devotha G. Nyambo

Glory C. Malamsha , Devotha G. Nyambo

©2023, GLORY C. MALAMSHA , DEVOTHA G. NYAMBO



This work is licensed under the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted use, distribution, and reproduction, provided the original work is properly credited. Cette œuvre est mise à disposition selon les termes de la licence Creative Commons Attribution (<https://creativecommons.org/licenses/by/4.0/legalcode>), qui permet l'utilisation, la distribution et la reproduction sans restriction, pourvu que le mérite de la création originale soit adéquatement reconnu.

IDRC GRANT / SUBVENTION DU CRDI : - CAPACITY STRENGTHENING IN RESPONSIBLE ARTIFICIAL INTELLIGENCE (AI) RESEARCH IN AFRICA: AI FOR DEVELOPMENT RESEARCH LAB IN ANGLOPHONE AFRICA

Multi-level Association Rule Mining for the Discovery of Strong Underrepresented Patterns

The Case Study of Small Dairy Farms in Tanzania

Glory C. Malamsha

Nelson Mandela African Institution of Science and Technology, Tanzania
malamshag@nm-aist.ac.tz
(corresponding author)

Devotha G. Nyambo

Nelson Mandela African Institution of Science and Technology, Tanzania
devotha.nyambo@nm-aist.ac.tz

Received: 16 January 2023 | Revised: 27 January 2023 | Accepted: 30 January 2023

ABSTRACT

Increasing the milk production of small dairy producers is necessary to cover the increase in milk demand in Tanzania. Currently, the population of people in both Tanzania and the world has increased and is predicted to increase more in the year 2050. The use of multilevel association rule mining methods to mine strong patterns among smallholder dairy farmers could help in identifying the best dairy farming practices and increase their milk production by adopting them. This study employed multi-level association rule mining to discover strong rules in three clusters, resulting in three levels of rules in each cluster. These three clusters were high, medium, and low milk producers. Rules were obtained for feeding practices, milk production, and breeding and health practices. These rules represent strong patterns among smallholder dairy farmers that could help them improve their dairy farming practices and have a gradual increase in milk production, from low to medium and from medium to higher milk production. Smallholder dairy producers would be provided with recommendations on their dairy farming practices, using rules based on the cluster to which they belong that could help them achieve higher milk production.

Keywords-association rules mining; dairy farming; smallholder dairy producers; milk yield

I. INTRODUCTION

Association rules are if-then statements that show the probability of relationships between data items or variables within large datasets [1, 2]. Association rule mining is used to discover interesting patterns between item sets in large databases [3-5]. The patterns discovered from association rule mining can be represented as rules, trees, or clusters which a user can interpret into knowledge [3]. Pattern matching is essential to identify relationships in datasets, and association rule extraction is among the most common techniques used in text mining [6]. Association rule mining provides association rules that exceed the user-specified minimum support and the minimum confidence level threshold [7].

A minimum of 145 days is required for small dairy farmers to learn farming strategies from other farmers in order to increase milk production [8]. This is a lot of time for a farmer who also needs to perform other activities. To minimize learning time, recommendations can be derived for small farmers using association rule mining methods. Association rule mining can mine data at single and multiple levels of abstraction depending on the approach used [9]. As sparse data

make it difficult to identify strong rules in low or original abstract layers, there is a need to mine strong rules from multiple abstract layers. Mining association rules in multiple abstract layers and formulate multilevel association rules is called multilevel association rule mining [10]. Recommendation systems use collaborative, content-based, and demographic filtering methods to search and filter large volumes of data and suggest items to users [11, 12]. The Syragri recommender system was created to provide recommendations to farmers on their practices. In [13], a hybrid approach was used for recommendation modeling to provide new items not rated by a user or rated by some users. As Indian farmers face various problems, such as insufficient harvesting and transportation methods, the government seeks ways to help them. The Collaborative Recommendation System for Agriculture Sector (CRSAS) was developed using English, Hindi, and Marathi to provide recommendations to farmers on various available agricultural programs and increase their knowledge of available government schemes [14]. Trustworthy recommendation systems are important and improve the ability of users to use them due to factors such as accuracy and good explanations [15].

The Tanzanian population was 34,443,603 in 2002, 44,928,923 in 2012, and was projected to increase to 59,441,988 in 2021 [16]. Currently, Tanzania's population increased to 61,741,120 people in 2022 [17], while the population worldwide is expected to increase to 8.9 billion people in 2050 from 7.9 billion in 2014 [18]. As the population and the per capita income increase, the demand for protein nutrients increases [19]. Ruminants, including cows, can be fed human-inedible plants such as alfalfa, pasture, and hay and produce high-quality protein products such as milk and meat [18]. These high-quality protein nutrient products would help meet the demands of an increasing population of people. Therefore, milk production as a protein source needs to be increased to meet the demands of the population. Small dairy farmers produce food for substantial populations of the world with small land holdings of less than 2 hectares [20]. Smallholder dairy farmers use low production costs, such as low-cost feeds and low technology, to produce milk which is cost-advantageous and helps solve the problem of food insecurity by providing milk proteins [21]. The challenges that small farmers face include poverty, food insecurity, lack of knowledge, lack of technical know-how, and poor support service [20, 21].

In [22], the K-Means and Self-Organizing Map (SOM) models were used to group small Tanzanian dairy farmers into six clusters from PEARL data. Farmer groups are a good tool to facilitate the development of small dairy farmers from subsistence to commercialization [22]. Multilevel association rule mining through the Apriori algorithm can be used to identify strong rules among various farmer groups that will ease the provision of recommendations to particular groups. Grouping farmers into appropriate farm groups, based on their characteristics, could simplify the use of interventions and strategies to improve milk production [22]. Recommendation systems have been currently used to develop models using big data to facilitate decision-making activities, including increasing farm yields [13]. Farm groups can be analyzed in more detail and hence, recommendations can be provided on how to increase milk yield. These recommendations would guide farmers to improve from low to medium and then achieve high milk production. Given these advancements, this study presents a multilevel association rule analysis as a data mining method to discover important underrepresented patterns in small dairy producer systems. Tanzania's dairy farmers were used as a case study but this method can be replicated in any country with similar characteristics.

II. METHODOLOGY

The extraction of rules in the dataset was performed using a multilevel association rule mining method with multiple levels of abstraction [9]. The application of multilevel mining of rules is increasing, as it is necessary to acquire precise information in various fields [23]. Many strong association rules are not obtained at a single level of abstraction [24]. Multi-level association rule mining requires viewing rules among dairy farmers at different multiple levels of abstraction. This study adopted a pre-clustered dataset with 6 dairy production clusters [22]. The 6 clusters were evaluated based on their dairy production performance using traits such as vaccination

frequency and milk production [25]. This study used 3 clusters to represent high, medium, and low dairy producers. High dairy producers were identified as cluster 1 (1180 households), medium dairy producers were identified as cluster 2 (516 households), and low dairy producers were identified as cluster 3 (295 households). In [25], cluster 1 had the best performance of 53%, while clusters 2 and 3 had 43% and 33% performance in product traits, respectively. The dataset consisted of 38 production traits that were represented with categorical data values. Table I shows the production traits, data types, and range of values of this dataset.

This study used the Apriori algorithm used to mine the association rules. This algorithm generates sets of frequent item sets in a database in iterations using a bottom-up approach by adding one item at a time [3]. The evaluation metrics used were the support, confidence, and lift values. Support and confidence are the two basic metrics in identifying association rules [2-4] while this study aimed to have a lift value greater than 1 to indicate stronger rules [7-26]. Support is defined as the proportion of records that contain $X \cup Y$ to the overall records in the database, whereas confidence is defined as the proportion of the number of transactions that contain $X \cup Y$ to the overall records that contain X [4]. Lift can be interpreted as the deviation of the support of the whole rule from the support expected under independence, given the support of both sides of the rule [7], while X and Y are antecedents and consequents of the extracted association rule, respectively.

Rule: $X \Rightarrow Y$

$$\text{support}(XY) = \frac{\text{support sum of } XY}{\text{Overall records in the database}} \quad (1)$$

$$\text{confidence} \left(\frac{X}{Y} \right) = \frac{\text{support}(XY)}{\text{support}(X)} \quad (2) [4]$$

$$\text{lift}(X \Rightarrow Y) = \frac{\text{support}(X \cup Y)}{(\text{support}(X) \text{ support}(Y))} \quad (3) [7]$$

Filtering of the rules was performed in various groups to obtain the important rules in each category with minimum support and minimum confidence for each cluster. Minimum support was selected based on the frequency of items in the various groups to extract the association rules [10]. Three levels of rules were obtained in each cluster. The dominating production traits were removed after levels 1 and 2 to obtain more rules with unique production traits in all clusters. Minimum support and confidence were also reduced at levels 2 and level 3 of all clusters to obtain unique rules.

III. RESULTS

Multilevel association analysis was performed for the 3 clusters. For each cluster, level 1 included rules on feeding and yield practices, level 2 included rules on health and breeding practices, and level 3 consisted of rules on breeding and yield practices.

A. Cluster 1: High Milk Producers

1) Level 1

Level 1 included 81 rules from the high milk-producing cluster, with minimum support of 0.6 and minimum confidence of 0.95, and lift range between 1.003 and 1.29. Strong rules

with lift values of 1.29, high confidence levels above 99%, and support range of 61-73% were found. The use of only stall-feeding systems in the rainy and dry seasons and bull breeding for the last calving had a high lift value of 1.29 with a support of 67%. Farmers owned 1 to 3 cows, they had frequent visits from 1 to 9 extension officers per year, no land for fodder production, no employees, and no training in dairy care at a

high lift value of 1.29 with support ranging from 61% to 69%. Other traits included farmers preferring bull breeding, 1 to 7 years of attending school, no fodder purchases, and no household members participating in the farmer groups, so the practice of farm activities usually depends on the basic knowledge of dairy farming.

TABLE I. PRODUCTION FEATURES USED IN MULTILEVEL ASSOCIATION RULE MINING

S/N	Production features	Data type	Range of values
1	Feeding system used during rainy season	Nominal	Mainly grazing, mainly stall, only ggazing, only stall feeding
2	Feeding system used during the dry	Nominal	Mainly grazing, mainly stall, only ggazing, only stall feeding
3	Reasons for choosing management (feeding) system	Nominal	Prevents diseases, cheap, extension officer advice, geographic conditions, insufficient land, large tracts of land, other reasons
4	Months which crop residue used for feeding	Discrete	1-12
5	Months which concentrates are used for feeding	Discrete	1-12
6	Months which fodder/feeds were purchased	Discrete	1-12
7	Watering frequency of the cattle	Discrete	1-4
8	Distance to water source (kn)	Interval	Less than 1, 1-3, 3-5, 6-10
9	Area under fodder production (acres)	Interval	Less than 0.5, 0.5-1, 1.1-2.5, 3.5-5, 5.5-10
10	Breed rank 1	Nominal	Ayrshire, cross, don't know, Friesian, Holstein, Jersey
11	Breed rank 2	Nominal	Ayrshire, cross, don't know, Friesian, Guernsey, Holstein, Jersey, other
12	Breed rank 3	Nominal	Ayrshire, cross, don't know, Friesian, Guernsey, Holstein, Jersey
13	Traits rank 1	Nominal	Milk quantity traits, body weight traits, calving traits, carcass traits, dairy type traits, disease traits, growth rate traits, milk feed traits, milk quality traits, reproductive traits, udder traits, other traits
14	Traits rank 2	Nominal	Milk quantity traits, body weight traits, calving traits, carcass traits, dairy type traits, disease traits, growth rate traits, milk feed traits, milk quality traits, reproductive traits, udder traits, other traits
15	Traits rank 3	Nominal	Milk quantity traits, body weight traits, calving traits, carcass traits, dairy type traits, disease traits, growth rate traits, milk feed traits, milk quality traits, reproductive traits, udder traits, other traits
16	Breeding method used during last calving	Boolean	Bull or artificial insemination
17	Preferred breeding method	Boolean	Bull or artificial insemination
18	Distance to breeding service provider	Interval	Less than 1, 1-5, 6-15, 16-30, above 30
19	Frequency of deworming cattle	Discrete	0-4
20	Self-deworming service	Boolean	Yes or No
21	Frequency of spraying cattle for pest control	Discrete	0-7
22	Frequency of vaccinating the cows	Discrete	0-4
23	Frequency of visit by extension officer	Interval	None, 1-9, 10-29, 30-49, 50 and above
24	Training	Nominal	No training, one day training, one week training, two weeks training, more than one month training
25	Experience in dairy farming	Discrete	3, 8, 13, 18, 23, 28, 35, 45, 50
26	Years of schooling of smallholder farmers	Interval	0, 1-7, 8-11, 12-16, above 16
27	Number of employees	Discrete	0-3
28	Household member in a farmer group	Nominal	Head household, son, spouse, daughter, other
29	Total land holding	Nominal	Average, below average, large
30	Total number of cattle	Interval	1-3, 4-6, 7-10, 11-16, 17-20, above 20
31	Number of livestock	Interval	1-20, 21-40, 41-60, 61-80, 81-100, hundreds, thousands
32	Preferred buyers	Nominal	Dairy chilling plants, hotels, individual consumers, milk collection center, other institutions, private milk traders
33	Distance to milk buyers	Interval	1-10, 11-20, above 20
34	Distance to market	Interval	Less than 1, 1-5, 6-10
35	Milk reserves for home consumption	Nominal	Above average, average, low
36	Liters of milk sold	Nominal	Above average, average, below average
37	Peak milk production of best cow	Nominal	Above average, average, below average
38	Number of milking cows	Interval	1-3, 4-8, 9-15

2) Level 2

At this level, the minimum support was adjusted to 0.22, the minimum confidence to 0.7, and the lift range was 1.02 to 1.43 for the 86 rules obtained. At lift values of 1.3, farmers dewormed their cattle 3 times a year, their distances to the market and the breeding provider were 1 to 5 kilometers, they sprayed their cattle 6 times a year, and did not self-deworm on

their cattle. At lift values greater than 1.2, the cattle vaccination frequency was twice a year, the liters of milk sold by farmers were average, farmers used concentrate for 3 months, and their owned total land was average. At 1.1 lift, the number of livestock was not identified, the preferred buyers were individual consumers, the watering frequency was twice a day, and the milk production of the best cow was average.

3) Level 3

Level 3 had 84 rules, where the minimum support was 0.164, the minimum confidence level was 0.67, and lift range was between 1.04 and 3.57. Strong rules were aligned from 80% to 100% confidence level with lift values greater than 1.9 and low support from 16.4%. These rules included preferred breed rank types and milk quantity traits in trait rank 1, while farmers owned 4 to 6 cattle and had 8 years of experience in dairy farming. The preferred breed rank types were identified at a lift of 3.5, where breed rank 1 was Friesian, breed rank 2 was Ayrshire, and breed rank 3 was Jersey. Some farmers had no preferred breed types in all 3 breed ranks but had eight years of experience in dairy farming at a lift of 2.0. Farmers preferred milk quality traits in trait rank 2, had average liters of milk sold, and animals were fed crop residue for 6 months with a lift value of 2.0. Some farmers also had 13 years of dairy farming experience and chose the feeding system for disease prevention.

B. Cluster 2: Medium Milk Producers

1) Level 1

In level 1, 0.4 minimum support and 0.86 minimum confidence level were used to obtain 100 rules from medium milk producers at a lift range between 1.07 and 1.39. Strong rules were located in two parts. The rules of the first part had a confidence level of 100%, 0.4-0.41 support, and 1.39 lift range. The preferred breeding method was artificial insemination at a confidence level of 100%, and farmers owned 1 to 3 milking cows. The rules of the second part had support and confidence levels between 0.4-0.42 and 0.87-0.91, respectively, with a lift range of 1.08 to 1.34. Farmers used artificial insemination in the last calving and were not members of farm groups at a lift value of 1.39. At lift values more than 1.3, the preferred buyers were individual consumers, the distance from the water source was less than 1km, and the watering frequency for the cattle was once per day. Farmers had average peak milk production for their best cows and the distance to the breeding service provider was 1-5km at a lift value of 1.1. At a 0.46 support, farmers received 1-9 visits from extension officers.

2) Level 2

The production traits that dominated level 1 were reduced in level 2. The minimum support used was 0.2, the minimum confidence level was 0.6, and 83 rules were obtained for a lift range between 1.02 and 3.15. Strong rules were determined for support between 0.22 and 0.35 and confidence levels from 0.85 to 1.0. Support was low but the confidence and the lift of the rules were high. Gaps in the 3 breed ranks were observed in both antecedents and consequents with a high support of 0.3 and lift of 3.0. However, at 0.25 support and lift above 3.0, the preferred breed in breed rank 1 was Friesian, breed rank 2 was Ayrshire, and breed rank 3 was Jersey. Farmers using Ayrshire and Friesian breeds performed deworming 3 times a year at a high lift above 2.0. Farmers did not purchase fodder and had low residue milk with a high lift of 1.5. Farmers with milk quantity traits in trait rank 1 performed deworming 3 times per year with a high lift above 2.0. Milk quality traits preferred in traits rank 1 and Friesian breed were observed with a high lift of 2.0.

3) Level 3

The production traits that dominated level 2 were reduced in level 3. A minimum support of 0.125 and a minimum confidence level of 0.6 were used in the mining process, the lift range obtained was between 1.01 and 2.66, and 70 rules were obtained for medium milk producers at this level. In these rules, most farmers preferred milk quality, milk quantity, and udder traits. Farmers who preferred milk quality traits in rank 2 had 1 to 20 cattle with a lift value of 2.0. In rank 1, farmers preferred milk quantity and had low residue milk for home consumption at a lift value of 2.0. Farmers did not have land for feed production, did not purchase feed, preferred udder traits in rank 3, and household members had attended school for 1-7 years at a lift value of 2.0. The preferred feature in rank 1 was the dairy type with a high lift of 2.5. At 1.5 lift, farmers purchased concentrates throughout the year and had a spraying frequency 2 per year. Other production traits included the use of crop residues for 4 months, farmers owned area for feeding production from 0.5 to 1 acres, and household members who had attended school for 8-11 years.

C. Cluster 3: Low Milk Producers

1) Level 1

At this level, the minimum support was 0.73, the minimum confidence level was 0.9, the lift range was between 1.004 and 1.14, and 76 rules were established. Farmers preferred and used the bull breeding method in the last calving and had no employees at a high support of 0.77 and lift above 1.1. Farmers did not have training in dairy farming at 0.75 support, did not carry out self-deworming at 0.75 support and above 1.1 lift, and did not have members in farmer groups at 0.74 support. Farmers used only stall feeding in both rainy and dry seasons, owned 1 to 3 milking cows, and had frequent visits, between 1 to 9, from extension officers at 0.74 support and 1.05 lift.

2) Level 2

The production traits that dominated level 1 were reduced in level 2. The minimum support was 0.34, the minimum confidence level was 0.72, the lift range was between 1.01 and 2.07, and 57 rules were obtained. Strong rules with a lift value of 2.07 were located at a confidence level of 1.0 and a support range from 0.36 to 0.39. No preferred breed type was identified at 2.0 lift and 0.37 support. The total land owned by farmers was below average and the distance to the market was 1-5km with 1.1 lift and 0.35 support. The vaccination frequency was twice a year and the watering frequency was once a day at 1.1 lift and 0.35 support. There was no distance to the buyer at 1.1 lift and 0.39 support. The liters of milk sold by the farmers were average and the household members had attended school for 1 to 7 years.

3) Level 3

The dominating production traits at level 2 were reduced in level 3. The used minimum support was 0.27, the minimum confidence level was 0.64, the lift range was between 1.03-1.85, and 51 rules were extracted. Farmers chose their feeding system due to insufficient land at 1.8 lift and preferred the trait of milk quality at 1.7 lift. Farmers performed spraying of cattle twice per year at 1.6 lift and deworming 3 times a year at 1.4

lift. At 1.2 lift, farmers used concentrates for 3 months, had low residue milk, their distance to the breeding service provider was 1-5km, and did not purchase fodders. Unique production traits included individual consumers as buyers and average milk production of the best cow.

Table II shows a summary of all the rules obtained in multilevel association rule mining from small dairy farmers in Tanzania.

TABLE II. SUMMARY OF RULES OBTAINED IN MULTILEVEL ASSOCIATION RULE MINING

Clusters	Levels	Number of rules	Minimum support	Minimum confidence level	Lift range
High producers (cluster 1)	Level 1	81	0.6	0.95	1.003 to 1.29
	Level 2	86	0.22	0.7	1.02 to 1.43
	Level 3	84	0.164	0.67	1.04 to 3.57
Medium producers (cluster 2)	Level 1	100	0.4	0.86	1.07 to 1.39
	Level 2	83	0.2	0.6	1.02 to 3.15
	Level 3	70	0.125	0.6	1.01 to 2.66
Low producers (cluster 3)	Level 1	76	0.73	0.9	1.004 to 1.14
	Level 2	57	0.34	0.72	1.01 to 2.07
	Level 3	51	0.27	0.64	1.03 to 1.85

IV. DISCUSSION

Most milk producers in all clusters practice only stall-feeding systems in both dry and rainy seasons. Feeding adequate meals to cattle with the required nutrients could help increase milk production. Milk production drops during the dry season for most farmers due to inappropriate methods to cope with it [21]. Apart from stall feeding, small dairy farmers provide crop residues and concentrates to their cattle. High milk producers provided crop residues to their cattle 6 months a year, medium milk producers provided crop residues for 4 months, while crop residues were not identified in small producers. The adoption of methods to produce and store crop residues is important and should be encouraged, including the treatment of crop residues for use during the dry season [21]. Farmers in high and low milk-producing clusters use concentrates for 3 months per year, but the medium milk-producing cluster has farmers who use concentrates for 3 to 12 months a year. Most farmers do not have or possess less than 1Km² of land for fodder production, which is insufficient for their cattle. In addition, farmers do not purchase fodder for their cattle. The results on the feeding systems highlight a major challenge towards higher productivity, as the practices are not in agreement with the available studies [21, 27-28]. With a zero-grazing set-up, a minimum of 1 acre of Napier grass per cow was emphasized to be adopted by small farmers as the primary source of fodder in Kenya [27]. Planting Napier grass on 1 acre with good management can help small farmers to feed 2 dairy cows during a year [28]. Effective feed production technologies have not been adopted due to cost, insufficient land, and high labor demands [21]. High milk producers have a higher watering frequency than the others, as they provide water at least twice a day.

Most farmers in high- and medium-producing clusters have 3 preferred breeds: Friesian, Ayrshire, and Jersey. Low milk producers did not have preferred breed types, raising questions about their dairy orientation and breeding selections [29]. The use of improved crossbreeding types is preferred to increase milk production compared to pure exotic breeds that have low production due to the production environments that are highly resource deprived. Improved breeds that are well suited for the environment of Tanzania, such as Friesian, Ayrshire, and

Jersey, have been proven to be among the top 10 effective in dairy production [29]. Farmers in high and low clusters preferred the use of bull breeding while farmers in the medium cluster preferred artificial insemination, which is an efficient and effective tool to increase milk production by breeding many cows with semen from good breeds in different geographic locations. Studies have indicated that artificial insemination is more effective than bull breeding because, in the former case, semen is better examined for diseases, quality, and fertility [30]. These results still yield uncovered service gaps for the farmers as they still rely on poorly performing breeding methods.

Most high milk-producing farmers prefer milk quantity, milk quality, and udder traits. Farmers in the medium-producing cluster prefer milk quality, milk quantity, dairy type, and udder traits. Farmers in the low-producing cluster prefer milk quantity and quality. In Tanzania, small dairy farmers prefer mostly cows with high milk production genetics but at an affordable price [31]. Farmers choose the best traits to increase milk production. This can be seen in high and medium producers who are much more aware of the good traits of their cattle than low producers.

Farmers in high and low milk production clusters carried out deworming of their cattle 3 times a year, but farmers in the medium cluster carried out deworming 3 to 4 times a year. Farmers in the high milk-production cluster spray their cattle 6 times a year, medium producers spray their cattle twice a year, and low-milk producers do not spray their cattle. The vaccination frequency in all clusters is twice a year. Most farmers in all clusters have 1 to 9 extension visits.

Insufficient and inadequate inputs such as feed, breed, health, and services such as lack of extension visits lead to the inability to control cattle diseases, poor livestock farming, and lack of knowledge and information [32]. Most of the farmers in all clusters did not perform cattle self-deworming, had not attended training, had no employees, and were not members of farmer groups. Farmers in the high cluster have more experience in dairy farming, from 8 to 13 years, than farmers in the medium and low clusters.

In terms of education level, farmers in high and low clusters attended school for 1 to 7 years, while farmers in the medium cluster attended school for 8 to 11 years. This highlights the gap between experienced farmers with low education levels against those who have significantly attained formal education. Although there were no significant differences in formal education between high and low producers, more research is needed to determine whether formal education can promote the adoption of improved technologies, such as the use of improved breeding methods [30]. The trained farmers were shown to have higher milk production than the others in Temeke [33]. Knowledge of how to conduct dairy farming practices is important to achieve high milk production. Farmers' knowledge in terms of receiving frequent extension visits has shown concern in the health of their animals; high milk producers were more experienced than the others, and hence had more knowledge in breeding practices.

Farmers in all clusters owned 1 to 3 milking cows with their best average production quantity of milk. The best cows from most small dairy farmers with high milk production in Tanzania produced 14.45 ± 5.12 L milk per day, the best cows from most small farmers with medium milk production small farmers produced 11.08 ± 4.29 L milk per day, and the best cows from low milk producers produced 9.15 ± 3.25 L milk per day. The cattle keepers in Tanzania sell small amounts of milk to consumers and have not met the demand for milk in a country where the population, income, and urbanization continue to increase [32]. The liters of milk sold by most farmers are average in all clusters. The preferred buyers in all clusters are individual consumers at distance ranging between 1 and 5 km. An increase in the quantity and quality of milk production can be achieved by an efficient milk collection system and the production of dairy products that could meet customer needs [34]. More farmers in the market will trigger more milk production since they will be confident in selling their products. Milk reserves for home consumption were low for the high and low clusters whereas the medium had a mix of farmers with average milk reserves.

The data mining process extracts useful information from large datasets that can uncover hidden patterns and is also called Knowledge Discovery or Knowledge Extraction [35]. Association rule mining enabled the mining of association rules from a dataset that identified underrepresented patterns among small farmers. This rule-based engine gives farmers a platform to learn various farming techniques from each other in their respective farm groups while extension officers can provide timely assistance [25]. Therefore, recommendations can be provided to the farmers according to their various practices.

V. CONCLUSION

It is necessary to provide farmers with knowledge about the production traits that can help them achieve higher milk production. This knowledge will not only increase production, but can also contribute to market products, employ personnel, and reduce poverty and food insecurity. Training small dairy farmers is also important with the available technology, such as feed technology, in a way that is easily accessible. The use of a rule-based engine to assign farmers to farm clusters enables

them to connect with the extension officers and enhance the information sharing and engagement. In a rule-based engine platform, farmers can receive recommendations based on the cluster to which they belong, as they are already available. A future study should investigate the use of these rules to provide recommendations to farmers based on these production features.

ACKNOWLEDGEMENT

The following agencies are acknowledged: Scholarship funders of the International Development Research Centre (IDRC) and Swedish International Development Cooperation Agency (SIDA) Scholarship program, Artificial Intelligence for Development (AI4D), Africa Scholarship Fund Manager, and Africa Center for Technology Studies (ACTS).

REFERENCES

- [1] A. S. Osman, "Data Mining Techniques: Review," *International Journal of Data Science Research*, vol. 2, no. 1, pp. 1–5, Jul. 2019.
- [2] T. A. Kumbhare and S. V. Chobe, "An Overview of Association Rule Mining Algorithms," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 1, pp. 927–930, 2014.
- [3] B. Kamsu-Foguem, F. Rigal, and F. Mauget, "Mining association rules for the quality improvement of the production process," *Expert Systems with Applications*, vol. 40, no. 4, pp. 1034–1045, Mar. 2013, <https://doi.org/10.1016/j.eswa.2012.08.039>.
- [4] T. Karthikeyan and N. Ravikumar, "A Survey on Association Rule Mining," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, no. 1, pp. 5223–5227, 2014.
- [5] A. Ksiksi and H. Amiri, "Using Association Rules to Enrich Arabic Ontology," *Engineering, Technology & Applied Science Research*, vol. 8, no. 3, pp. 2914–2918, Jun. 2018, <https://doi.org/10.48084/etasr.1998>.
- [6] A. Alqahtani, H. Alhakami, T. Alsubait, and A. Baz, "A Survey of Text Matching Techniques," *Engineering, Technology & Applied Science Research*, vol. 11, no. 1, pp. 6656–6661, Feb. 2021, <https://doi.org/10.48084/etasr.3968>.
- [7] M. Hahsler, "arulesViz: Interactive Visualization of Association Rules with R," *The R Journal*, vol. 9, no. 2, pp. 163–175, 2017, <https://doi.org/10.32614/RJ-2017-047>.
- [8] D. G. Nyambo, E. T. Luhanga, Z. O. Yonah, F. D. Mujibi, and T. Clemen, "Leveraging peer-to-peer farmer learning to facilitate better strategies in smallholder dairy husbandry," *Adaptive Behavior*, vol. 30, no. 1, pp. 51–62, Feb. 2022, <https://doi.org/10.1177/1059712320971369>.
- [9] P. Manda, S. Ozkan, H. Wang, F. McCarthy, and S. M. Bridges, "Cross-Ontology Multi-level Association Rule Mining in the Gene Ontology," *PLOS ONE*, vol. 7, no. 10, 2012, Art. no. e47411, <https://doi.org/10.1371/journal.pone.0047411>.
- [10] J. Tan, "Different Types of Association Rules Mining Review," *Applied Mechanics and Materials*, vol. 241–244, pp. 1589–1592, 2013, <https://doi.org/10.4028/www.scientific.net/AMM.241-244.1589>.
- [11] P. K. Biswas and S. Liu, "A hybrid recommender system for recommending smartphones to prospective customers," *Expert Systems with Applications*, vol. 208, Dec. 2022, Art. no. 118058, <https://doi.org/10.1016/j.eswa.2022.118058>.
- [12] V. Rohilla, M. Kaur, and S. Chakraborty, "An Empirical Framework for Recommendation-based Location Services Using Deep Learning," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9186–9191, Oct. 2022, <https://doi.org/10.48084/etasr.5126>.
- [13] J. Konaté, A. G. Diarra, S. O. Diarra, and A. Diallo, "SyrAgri: A Recommender System for Agriculture in Mali," *Information*, vol. 11, no. 12, Dec. 2020, Art. no. 561, <https://doi.org/10.3390/info11120561>.
- [14] S. Jaiswal, T. Kharade, N. Kotambe, and S. Shinde, "Collaborative Recommendation System For Agriculture Sector," *ITM Web of Conferences*, vol. 32, 2020, Art. no. 03034, <https://doi.org/10.1051/itmconf/20203203034>.

- [15] M. Dong, F. Yuan, L. Yao, X. Wang, X. Xu, and L. Zhu, "A survey for trust-aware recommender systems: A deep learning perspective," *Knowledge-Based Systems*, vol. 249, Aug. 2022, Art. no. 108954, <https://doi.org/10.1016/j.knosys.2022.108954>.
- [16] "2021 Tanzania in Figures," National Bureau of Statistics, Dodoma, Tanzania, Jun. 2022.
- [17] "Census Information Dissemination Platform." <https://sensa.nbs.go.tz/>.
- [18] G. Wu *et al.*, "Production and supply of high-quality food protein for human consumption: sustainability, challenges, and innovations," *Annals of the New York Academy of Sciences*, vol. 1321, no. 1, pp. 1–19, 2014, <https://doi.org/10.1111/nyas.12500>.
- [19] M. J. Boland *et al.*, "The future supply of animal-derived protein for human consumption," *Trends in Food Science & Technology*, vol. 29, no. 1, pp. 62–73, Jan. 2013, <https://doi.org/10.1016/j.tifs.2012.07.002>.
- [20] G. Rapsomanikis, "The economic lives of smallholder farmers: An analysis based on household data from nine countries," Food and Agriculture Organization of the United Nations, Rome, 2015.
- [21] D. Maleko, G. Msalya, A. Mwilawa, L. Pasape, and K. Mtei, "Smallholder dairy cattle feeding technologies and practices in Tanzania: failures, successes, challenges and prospects for sustainability," *International Journal of Agricultural Sustainability*, vol. 16, no. 2, pp. 201–213, Mar. 2018, <https://doi.org/10.1080/14735903.2018.1440474>.
- [22] D. G. Nyambo, E. T. Luhanga, Z. O. Yonah, and F. D. N. Mujibi, "Application of Multiple Unsupervised Models to Validate Clusters Robustness in Characterizing Smallholder Dairy Farmers," *The Scientific World Journal*, vol. 2019, Jan. 2019, Art. no. e1020521, <https://doi.org/10.1155/2019/1020521>.
- [23] R. Yadav, "An Improved Multiple-level Association Rule Mining Algorithm with Boolean Transposed Database," *International Journal of Computer Science and Information Security*, vol. 13, no. 8, pp. 95–104, 2015.
- [24] Y. Cheng, W. Yu, and Q. Li, "GA-based multi-level association rule mining approach for defect analysis in the construction industry," *Automation in Construction*, vol. 51, pp. 78–91, Mar. 2015, <https://doi.org/10.1016/j.autcon.2014.12.016>.
- [25] F. Mavura, S. M. Pandhare, E. Mkoba, and D. G. Nyambo, "Rule-Based Engine for Automatic Allocation of Smallholder Dairy Producers in Preidentified Production Clusters," *The Scientific World Journal*, vol. 2022, Jun. 2022, Art. no. e6944151, <https://doi.org/10.1155/2022/6944151>.
- [26] D. G. Nyambo, E. T. Luhanga, and Z. O. Yonah, "Characteristics of smallholder dairy farms by association rules mining based on apriori algorithm," *International Journal of Society Systems Science*, vol. 11, no. 2, pp. 99–118, Jan. 2019, <https://doi.org/10.1504/IJSSS.2019.100101>.
- [27] L. M. Mburu, "Effect of seasonality of feed resources on dairy cattle production in coastal lowlands of Kenya," Ph.D. dissertation, University of Nairobi, Kenya, 2015.
- [28] "Hybrid Napier Grass Cultivation For Dairy Animals | Agri Farming," Mar. 07, 2019. <https://www.agrifarming.in/hybrid-napier-grass-cultivation-for-dairy-animals>.
- [29] F. D. N. Mujibi *et al.*, "Performance Evaluation of Highly Admixed Tanzanian Smallholder Dairy Cattle Using SNP Derived Kinship Matrix," *Frontiers in Genetics*, vol. 10, 2019.
- [30] S. Patil, S. Karunamay, and S. Nath, "Importance of Artificial Insemination in Dairy Farming –," *Pashudhan praharee*, May 31, 2020. <https://www.pashudhanpraharee.com/importance-of-artificial-insemination-in-dairy-farming-4/>.
- [31] A. R. Chawala, G. Banos, A. Peters, and M. G. G. Chagunda, "Farmer-preferred traits in smallholder dairy farming systems in Tanzania," *Tropical Animal Health and Production*, vol. 51, no. 6, pp. 1337–1344, Jul. 2019, <https://doi.org/10.1007/s11250-018-01796-9>.
- [32] A. Omere, M. Kidoido, E. Twine, L. Kurwijila, M. O'Flynn, and J. Githinji, "Using 'theory of change' to improve agricultural research: recent experience from Tanzania," *Development in Practice*, vol. 29, no. 7, pp. 898–911, Oct. 2019, <https://doi.org/10.1080/09614524.2019.1641182>.
- [33] K. S. Chausa, "Assessment of factors affecting performance of dairy cattle kept in smallholder farms in peri-urban areas of Temeke municipality," Ph.D. dissertation, Sokoine University of Agriculture, Morogoro, Tanzania, 2022.
- [34] E. Ulicky, J. Magoma, H. Usiri, and A. Edward, "Improving Smallholder Livelihoods: Dairy Production in Tanzania," *International Grassland Congress Proceedings*, Apr. 2020.
- [35] A. H. Blasi, M. A. Abbadi, and R. Al-Huweimel, "Machine Learning Approach for an Automatic Irrigation System in Southern Jordan Valley," *Engineering, Technology & Applied Science Research*, vol. 11, no. 1, pp. 6609–6613, Feb. 2021, <https://doi.org/10.48084/etasr.3944>.