Agent Mining Framework for Analyzing Moroccan Olive Oil Datasets

Belabed Imane[1], Jaara El Miloud[2], Belabed Abdelmajid[3], Talibi Alaoui Mohammed[4]

University Mohammed the First, Oujda, Morocco[1, 2, 3]

Faculty of Science and Technology, Fez, Morocco[4]

Abstract—**Data mining and intelligent agents have become two promising research areas. Each intelligent agent functions independently while cooperating with other agents, to perform effectively assigned tasks. The main goal of this research, is to provide a mining implementation that can help biological researchers for discovering parameters that affect the cost of olive oil in Morocco. To solve this problem, we used a method involving two data mining techniques, clustering of variables, quantitative association rules and multi-agent system to fuse these two techniques. Therefore, we have developed a multi-agent framework that has been validated by using concrete data from the Provincial Direction of Agriculture of Berkane, Morocco. To prove the performance of our framework, we tested the proposed multi-agent tool using three datasets from different fields. Conforming to biological researchers, our method generates a clear knowledge because the framework proposes high-confidence rules that can correctly identify olive oil factors.**

Keywords—*Quantitative association rules; clustering of variables; multi-agent system*

I. INTRODUCTION

Data mining technology aims to find useful knowledge from the database. In fact, data mining technology engages a crucial role in many business analysis and prediction applications used to complete data analysis.

The clustering covers many various algorithms and methods for grouping similar kinds of objects into various categories. Such algorithms or methods are associated with organizing the observed data into expressive structures. If two objects belong to the same group the correlation between them is the largest, otherwise it is the smallest. In light of the above, cluster analysis can be used to discover structures in data without providing explanations. This research takes the partition based clustering such K-Means algorithm for variables.

The issue of finding relevant relationships between attributes has considerably studied through the association rules concept. Association rules are a technique that allows users to discover associations between different objects in databases. The results are given in the form of antecedent and consequent. The reliability of rules is usually measured by using statistical function, as for instance the support and the confidence. The rules that maximize the support and the confidence are more considered in the mining tasks.

The execution of association rules on numeric attributes, is problematic, as a conventional approach, the most methods perform a pre-processing phase called discretization, which refers to the process through which we can transform continuous variables, into a discrete form before executing the learning task.

To address this issue, in this work, we recommend the use of genetic algorithm to process quantitative datasets. Consequently exceed the discretization phase. The adoption of a genetic algorithm is sustained by our previous work, in which we evaluate the association rules obtained by running apriori and genetic algorithm using quantitative datasets in multi-agent environment. The experiments show that genetic algorithm avoids the redundant rules with satisfied execution time [22].

The most motivation of this research is to obtain an automatic solution that consolidates the two data mining techniques illustrated before. Therefore, the idea of integrating multi-agent technology into data mining applications turns to be useful. In fact, it seeks supplementary benefits by integrating agent technology into data mining technology. This interaction allows us to create an effective solution to solve the main problems of this work, which includes the discovery of factors that increase the profit of Moroccan olive oil and thus reduce the cost of olive oil.

The rest of the paper is split as follows. Section II discusses the data mining process adopted to implement the agent framework, while Section III briefly presents the data mining techniques used. Section IV presents an overview of multi-agent system concept. Section V presents the works related to our research area. Section VI details the architecture of the agent-mining tool, and presents the implementation process. Section VII illustrates the results obtained by applying the developed tool on three test cases specified in Table I. Section VIII outlines the use of the framework in the Moroccan agricultural domain test case. Section IX ends the paper and highlights suggestions for future work.

A. Motivation and Our Contribution

Over the last years, Morocco has developed a new agricultural strategy called Green Morocco Plan. The project was aimed at supporting modern agriculture, with benefit and high productivity that correspond to market requirements. The plan encourages private investments, in order to improve the chain of productivity and develop industrial activities related to agriculture.

The Green Morocco Plan is an agricultural strategy launched in 2008 that aims to make agriculture, the main growth engine of the national economy over the next ten to fifteen years, with significant benefits in terms of growth. New

instruments were designed for the Green Moroccan Plan implementation. The plan is structured around seven pillars. Most actions come under Pillar I and Pillar II, the goal of Pillar I of the Green Morocco Plan is the development of agriculture with high benefit and high productivity. This requires the voluntarism creation of agricultural development poles insuring high benefit and market requirements [2].

The pillar I aims at improving the production chain of high benefit crops such olive [1].

The goal of this paper is to deal with Pillar I, particularly the identification of the factors that optimize the Moroccan olive oil cost. Our research consists on creating an efficient process to discover theses parameters and support decision making in this field.

II.   RELATED WORK

In this section, we will briefly introduce previous researches in machine learning area including knowledge discovery in agent systems or multi-agent systems. The techniques used in these studies include association rule mining, clustering mining, and rule generation algorithms. The proposed approach is mainly related to two areas of research, knowledge extraction from dataset and knowledge modelling using the multi intelligent agent system. Popa proposed an intelligent recommendation system based on multi-agent, called Agent Discover. The purpose of the system is to solve the complexity of knowledge discovery and provide a tool to support researchers and non-expert users to explore knowledge discovery methods and quickly find results in the field [14]. Tong, proposed a real time Data Mining and Multi-Agent System called DMMAS. The DMMAS method uses data partitioning and multiple agents, and can choose to use heterogeneous or homogeneous data mining technology. Agent-based distributed processing can model and combine the results of all agents to improve the efficiency [15]. Nahar presented a paper in which she executes association rule mining and a computational intelligence to recognize the factors, which contributes to the apparition of heart diseases for males and females. This study proposed an experiment using three rule generation algorithms Apriori, Predictive Apriori and Tertius to extract rules from heart disease data [16].

Ait-Mlouk proposed a method based on multi-criteria analysis to discover a category of relevant association rules. The author uses multi-agent system to integrate, manage, and model the quality measurement according to six agents working in cooperation [17]. Kaur produced a spatial data mining techniques to extract implicit knowledge from spatial attributes. These techniques are applied in different fields such as healthcare, marketing, and remote sensing databases to improve planning and decision-making process [18].

Salleb-Aouissi, proposes a system based on generating association rules from quantitative datasets by using the concept of genetic algorithm. They tested their tool on both real datasets from medical domain and synthetic datasets [21].

The approach presented in this work, consists on providing an agent mining tool that includes two data mining techniques represented by two categories of mining agents. Thanks to the integration of genetic algorithm, the present solution uses one agent in the association rules phase. Thus, processes efficiently the issue of redundant rules presented in Ait-Mlouk [17]. In fact, in that work the authors use a multi criteria approach to filter significant rules. Consequently, they use six agents to generate association rules.

In addition, our proposed agent tool overcomes the Salleb-Aouissi approach mainly in the execution time. Due to the integration of both multi-agent system and K-Means clustering, the global execution time of every experiment is fourteen times lower than the best execution time in that work, for real datasets [21].

III.   DATA MINING ANALYSIS PROCESS

When we studied the problem of extracting knowledge from the Moroccan olive dataset, we request a solution based on the combination of clustering and association rules [3]. We recommend applying rule association algorithms to each of the defined clusters.

To achieve this goal, we choose K-Means for the clustering phase and genetic algorithm for association rules phase.

Among the knowledge discovery process, we confront two challenges such:

- Executing K-Means algorithm for variables.

- Extracting association rules from quantitative data.

To exceed the first constraint, we choose to process only quantitative datasets and transpose the data in order to cluster variables instead of observations. Then, we integrate K-Means algorithm using the Weka open source framework [4]. For the association rules phase, we chose to use genetic algorithm for quantitative association rules. The algorithm relies on genetic algorithm to find dynamically the optimal interval for numerical data.

The data used in this research are divided into two parts. The first one consists of using datasets from internet, uci-archives for machine learning [20], in order to validate the framework and evaluate the performance. The second part consists of analysing olive oil datasets, based on our framework. Table I, presents the datasets used in our experiments.

TABLE I.   DATA SPECIFICATION

Data field	Nature of variable	Volume of variables
Absenteeism data	Quantitative	19 variables
Bio-degradation data	Quantitative	41 variables
Air quality data	Quantitative	12 variables
Olive crops datasets	Quantitative	17 variables

IV. DATA MINING ALGORITHMS

In this section, we highlight the data mining techniques used.

A. K-Means Clustering

The K-Means algorithm based on division is a popular clustering algorithm. This unsupervised algorithm is used commonly, for data mining and pattern identification. The algorithm aims at grouping data of more or less similarity by finding iteratively the centroid between the elements. In this work, the distance metric adopted is the Euclidian distance. The algorithm works as follow:

Step 1: Select a number of clusters, k.

Step 2: Choose k, the initial starting values to be the initial centroids.

Step 3: Affect each point to the cluster whose centroid is nearest to it in term of Euclidian distance.

Step 4: When each point is assigned to a cluster, recalculate the new k centroids.

Step 5: Repeat steps 3 and 4 until no point changes its cluster assignment, or until a maximum number of iterations is reached.

B. Association Rules

Association rule is a frequent data mining technique usually used for discovering relevant relationships between variables in large databases [5]. It studies the frequency of items that appear together in the transaction database and identifies frequent item sets based on the first threshold called support. The second threshold is called confidence, which calculates the conditional probability that an item shows up in a transaction when another item appears. The form of the association rules is: A B [s, c], where A and B are conjunctions of attribute value-pairs, and s, for support, is the probability that A and B show up together in a transaction and c, for confidence, is the conditional probability that B appears in a transaction when A exists.

C. Genetic Algorithm Overview

Genetic algorithm GA is an adaptive method that can be used to solve search and optimization problems, which is based on biological genetic processes.

The algorithm is based on the genetic processes of biological organisms. Over many generations, natural populations evolve confirming to the principles of natural selection and survival of the fittest first. By imitating this process, genetic algorithms can evolve solutions to real-world problems. The genetic algorithm uses multiples solutions collectively known as population. These solutions are usually coded in binary strings. Every solution or individual is assigned a fitness, which matches with the objective function of the search. Thereafter, the individual populations were modified to new populations by applying three operators like natural genetic operators, namely reproduction, crossover, and mutation.

Nevertheless, genetic algorithms have at common the following elements: population, selection according to fitness, crossover to reproduce new offspring, and random mutation of new offspring.

Initial population: The generation of initial population is executed as follow: Firstly, we consider the interval $[a_i, b_i]$, which represents the domain of quantitative variable a^{th}. Then the length of interval is decreasing until attaining a minimum support specified by the user. We note that the bounds a_i and b_i are chosen at random. This enables starting with enough diversity within the initial population.

The mutation operator sustains the diversity within the population. Indeed, it uses a selected rate, which determines the degree of changes to be made. The change is performed in term of modifying the length of the interval of selected variable [6].

The crossover operator is equivalent to reproduction in biological crossover. It is founded on taking two individuals, called parents [6], and generating new individuals. In our case, the concept is applicable to variable. Thereby, the interval of each variable is either inherited from one parent or formed by merging the bounds of the two parents [7].

Fitness function: A fitness function is a type of objective function that determines the optimally of chromosome in traditional genetic algorithm [8]. In this paper, we filter rules by measuring the support and the confidence. Therefore, the fitness function is built by combining the support and confidence parameters. We predefine the thresholds of minimum support, min_sup, and the minimum confidence, min_conf, for the algorithm [9].

D. Algorithm

The algorithm selects high confidence rules. In fact, the algorithm starts with a set of predefined rules in term of the implication direction. In this work, both the condition and the conclusion parts of a rule are recognized. In fact, the variables that constitute the conditions are defined in the clustering phase. Additionally, for every dataset, we choose only one variable to represent the conclusion part. The characteristics of these variables are illustrated in Table II.

The algorithm finds the optimal interval for quantitative variables present in that rule template. An optimization criterion is applied to select only the rules that maximize the support and the confidence parameters.

TABLE II.   VARIABLES TO PREDICT

Data field	Variable to predict
Absenteeism data	Absenteeism hours
Bio-degradation data	Number of Halogen atoms
Air quality data	Air humidity
Olive crops data	Olive oil cost.

The algorithm follows the prototype of traditional genetic algorithm.

The algorithm inputs are the minimum confidence (min_conf), the minimum support (min_sup), the population size (Popsize), the crossover rate (Cross), the number of generations (Genum) and the mutation rates (Mut).

Genetic algorithm pseudo-code:

Select a set of attributes

Let *Rt* a set of predefined rule specified on selected variables

For each *r* ∈ *Rt* do

Choose the initial population P of Popsize

While *i* ≤ *Genum* do

Breed new generation through mutation and crossover operators i.e. Mut and Cross.

Extract the itemsets that deliver the best fitness to form the association rule values

i++

Return R= *max (*fitness (r)); r belongs to P.

V. AGENT AND MULTI-AGENT SYSTEMS

Various definitions have been proposed for the concept of multi-agent system (MAS). A multi-agent system is an approximately coupled network of problem-solver entities that collaborates to answer the issues related to the individual capabilities or knowledge of each entity.

Although the agents in a multi-agent system are often equipped with behaviors designed in advance, they often need to apprentice new behaviors online in order to improve progressively the performance of agents; consequently, the entire multi-agent system is ameliorated.

One of the current sectors of popularity of multi-agent systems is the smart agricultural domain especially the concept of internet of things where agents interact with each other to achieve their individual or shared goals. To perform in such an interactive environment, agents must surpass two challenges: they have to be ready to localize each other, since agents might appear, disappear, or change their status at any time. Additionally, the agents must be able to interact with each other [10].

A. Agents

An agent is defined as a computer system located in some environment and capable of acting autonomously in this environment to achieve its design goals. They can be reactive to changes produced in their environment; also, they are able to communicate and use computational intelligence to reach their goals by being proactive [11].

B. Multi-Agent Systems

By associating varied agents in one system to solve a problem, the produced system is called multi-agent system.

These systems hold multiple agents that individually solve problems. They can communicate with each other and assist each other in achieving larger and more complex goals [12]. Multi-agent systems have been used in predicting the stock market, industrial automation etc. In this work, we developed a multi-agent system for predicting the factors that optimize the olive oil cost.

C. Multi-Agent Systems and Data Mining

Data mining and multi-agent systems present attractive features to form systems that are more intelligent [13].

- The combination of multi-agent autonomy and knowledge data mining provides adaptable systems.

- Data mining techniques such as association rule extraction have no equivalent in agent systems. Currently, these techniques deliver agents with the ability of learning and discovering.

- Data mining can enhance the agent capability of handling uncertainty via historical event analysis, dynamic mining, and active learning. By mining agent behavioural data, it is possible to reach a balance between agent autonomy and adequate learning.

VI. MULTI-AGENT SYSTEM FOR DATA MINING SYSTEM

A. Architecture

As described below, the proposed mining framework shown in Fig. 1 includes four types of agents:

1) User agent.
2) Data agent.
3) Coordinator agent.
4) Clustering agent.
5) Association rules agent.

The agents are distributed in set of various containers taken from JADE, Java Agent Development Framework, in which our framework is implemented. One of these containers is the main container, which holds an AMS, Agent Management System and a DF, directory facilitator. The AMS agent is employed to manage the life cycle of other agents in the platform, and therefore the DF agent provides agent search services, such as yellow pages.

Fig. 1 presents a view of the framework architecture implemented in JADE. It shows the various categories of agents and their interactions along with the data mining algorithms. The figure illustrates the main container that holds the coordinator agent, additionally to the AMS agent and the DF agent as illustrated before. The other containers hold the specialized agent, the clustering agent, the data agent, the user agent and the association rules agent.
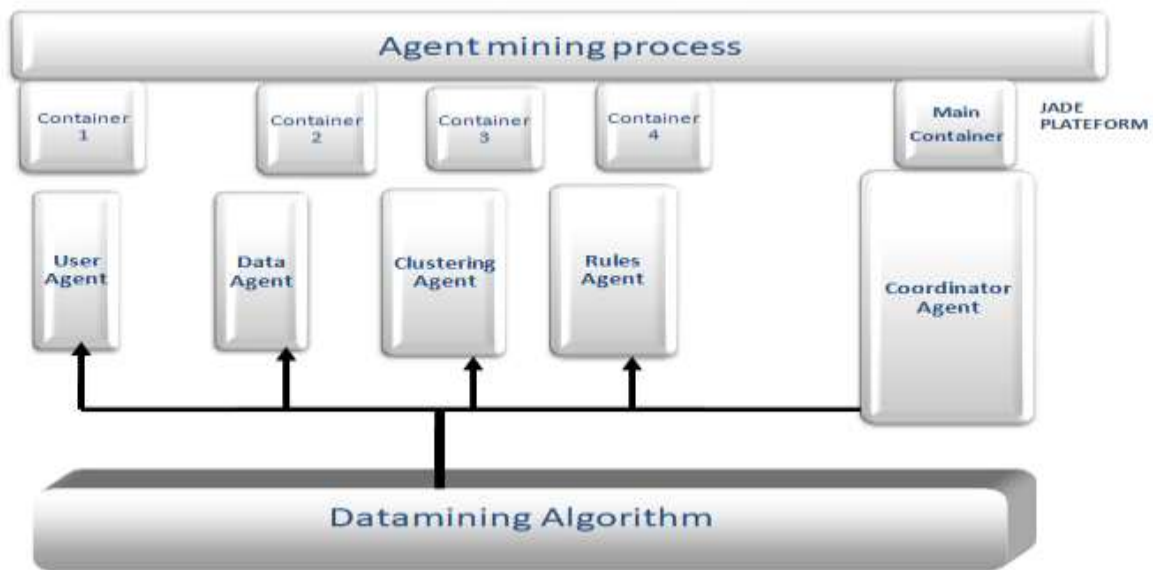
Fig. 1.   Framework Architecture.

As shown in Fig. 1, the agent mining process is composed of multiple collaborative agents that act in several levels.

- **User agent** constitutes the interface between end user and the mining framework. The agent is responsible for obtaining the K value of K-Means and the datasets path.

- **Data agent** is responsible for charging the datasets from the data source and keeping meta-data information about the data source. There is a direct liaison between a data agent and a selected data source. Data agent is in charge of forwarding their data, when requested.

- **Coordinator agent** ensures the proper transmission of messages among the agents. It collects the user specifications and sends them to the corresponding agent.

- **Clustering agent** executes K-Means algorithm. Once the clustering agent have complete it task, it informs the coordinator agent.

- **Association rules agent** is responsible for generating supervised rules inside each predefined cluster through the genetic algorithm.

VII.   Multi-Agent Framework Implementation

The present work proposes the implementation of the multi-agent system for data mining framework including clustering and association rules. We have developed java platform through Agent-Oriented Programming paradigm, AOP. The communication inter-agent is maintained through the recommendations of the standard Foundation for Intelligent, Physical Agents, FIPA. We have implemented our proposed tool with JADE [19].

JADE is FIPA-compliant middleware that enables the development of distributed applications based on the agent paradigm and is adequate to process large amounts of data with a data mining approach. JADE allows portability, which is assured by the use of Java, and defines an agent platform comprising a set of containers that may be distributed across a network. In the JADE platform, the main container holds a number of mandatory agent services, such as the AMS and DF agents. The DF agent is responsible for the yellow pages service at the main container of the JADE framework. In addition, the AMS agent is used to control the life cycles of other agents in the platform.

JADE holds two main products, a FIPA-compliant agent platform and a package to develop Java agents. JADE also provides the implementation of FIPA Agent Communication Language, ACL, which is a message-based protocol defined by FIPA.

Because this implementation requires a data-mining algorithm, the first implementation measure is to integrate the data-mining algorithm into the JADE platform. As a data mining tool, we choose the Weka 3.6.1 Java Library in the first phase of K-Means clustering.

VIII.   Experimental Results

In this section, we expose the ability of our framework to discover knowledge with test cases in different domains, particularly in environment and agriculture. We will focus on the advantage provided by multi-agent technology to promote the data mining performance. Thus, for every phase of our data mining process we will present the results obtained by executing the three datasets illustrated before in Table I. Concretely, the performance evaluation during the data mining process, is measured through the following metrics:

- *Memory used:* physical memory consumed by the task of the agent when it has been executed. The resulting value is given in megabytes (MB).

- *The execution time with agent:* The amount of time the task took to process by the agent. The resulting value is given in seconds (s).

A. *Clustering Phase*

Clustering is conducted with an experimental K value of K-Means set into three. Then the framework was executed with two modes: with agent and without it, in order to evaluate the impact of multi-agent implementation. The number of clusters and the number of association rules agents vary proportionally. In fact, if the number of clusters is significant, more than one association rules agent is needed. However, the proposed architecture is extensible to include high number of clusters if the number of association rules agents is increased. In that case, several agents should be added such, a new manager agent that should firstly, orchestrates the interaction inside the association rules agents groups and secondly manage the communication with the coordinator agent.

Table III presents the result of applying the multi-agent concept and it consequent performance in the clustering phase. Due to the integration of multi-agent concept, the execution time is reduced greatly. However, in term of memory consumption, the agent platform consumes more than the java environment. The increase of the memory consumption is due to the DF agent hold in the main container. In fact, the DF agent should store it catalogue in memory. It contains, the yellow pages, which include the matches between the description of the agents and their proposed services.

B. *Association Rules Phase*

As discussed in Section I. The process of data mining used in this work, focuses on extracting rules between each discovered cluster and the variables to predict illustrated in Table II. This concept is valid for the three data source, bio-degradation data, absenteeism data, and air quality data.

The experiments are conducted with the genetic algorithm using quantitative datasets. The experimental thresholds of genetic algorithm are set to be as follows:

- The minimum confidence threshold fixed to 60%.
- The minimum support threshold fixed to 10%.
- The population size fixed to 250.
- The crossover rate fixed to 50%.
- The mutation rate fixed to 40%.

We chose to start the genetic algorithm with relatively low thresholds in order to generate more rules. Consequently, we can evaluate correctly the robustness of our tool.

The results presented below, concern the performance of the multi-agent approach integration, for rules mining phase using genetic algorithm. Concretely, the next subsections illustrate the process of rules extractions with two different modes, with agent and without. Note that, the number of rules extracted depends only on the genetic algorithm and not on the integration of the multi-agent technology.

1) Biodegradation datasets: Table IV presents the results of the association rules phase. For bio-degradation datasets, the rules extracted with genetic algorithm in each cluster were respectively 80 rules in cluster 1, 79 rules in cluster 2 and one rule in cluster 3. This experiment presents a few runtime improvements with the agent integration approach. However, the multi-agent platform is memory intensive compared to the java environment.

2) Air quality datasets: Table V presents, the three cluster results, with their execution time and memory consumption. The performance runtime is increased slightly in cluster 1 and cluster 3. However, for the air quality datasets, the multi-agent platform is more memory consuming than the traditional approach. The rules extracted with genetic algorithm in each cluster were respectively 42 rules in cluster 1, 15 rules in cluster 2 and 42 rule in cluster 3.

3) Absenteeism datasets: Table VI illustrates, the generation of rules using genetic algorithm for the absenteeism data. The multi-agent platform presents consistent improvement of the execution time, especially in cluster 1. In fact, the execution time is fifteen times less than the traditional approach. Nevertheless, the multi-agent platform stills memory intensive.

In conclusion, of this part, Tables IV, V and VI present the three experiment tests included execution time, memory consumption and number of generated rules.

In the association rules phase, the results obtained by integrating the agent approach on the three datasets, show an insignificant improvement in the execution time, except for the absenteeism test case, cluster 1, where the multi-agent approach presents an interesting runtime improvements. We note that, in this work the number of clusters is approximately small. The benefit of the multi-agent approach will be significant in the case of large number of clusters. In that case, the sum of the execution times obtained in each cluster would be considerable. Nevertheless, in this phase the increase of the memory consumption is always present.

TABLE III. RESULT OF CLUSTERING PHASE

K=3	Data mining process clustering phase		
Data	**Volume**	**Clustering with agent**	**Clustering without agent**
Bio-degradation	41 variables	Time : $3.23*10^{-4}$ seconds Memory : 23.92MB	Time : 0.81 seconds Memory : 19,98 MB
Absenteeism	19 variables	Time: $3.67*10^{-4}$ seconds. Memory: 25.36 MB.	Time : 0.66 seconds Memory :20.20 MB
Air quality	12 variables	Time : $3.24*10^{-4}$ seconds Memory :22.54 MB	Time : 0.68 seconds Memory :19.35 MB

TABLE IV.   ASSOCIATION RULES RESULT FOR BIO-DEGRADATION DATA

Genetic algorithm for quantitative variables		Association rules phase		
		Association rules with agent	Association rules without agent	Number of rules extracted
Bio-degradation data	Cluster 1	Time : 0.01 seconds Memory : 23.49 MB	Time : 0.02 seconds Memory : 2.09 MB	80
	Cluster2	Time : 0.01 seconds Memory : 32.01 MB	Time : 0.02 seconds Memory : 18.55 MB	79
	Cluster 3	Time : 0.01 seconds. Memory : 33.67 MB.	Time : 0.02 seconds. Memory : 15.16 MB.	1

TABLE V.   ASSOCIATION RULES RESULT FOR AIR QUALITY DATASETS

Genetic algorithm for quantitative variables		Association rules phase		
		Association rules with agent	Association rules without agent	Number of rules extracted
Air quality data	Cluster 1	Time : 0.01 seconds Memory : 34.99 MB	Time : 0.02 seconds Memory : 25.03 MB	42
	Cluster2	Time : 0.02 seconds Memory : 30.09 MB	Time : 0.02 seconds Memory : 23.49 MB	15
	Cluster 3	Time : 0.01 seconds. Memory : 36.10 MB.	Time : 0.02 seconds. Memory : 23.83 MB.	42

TABLE VI.   ASSOCIATION RULES RESULT FOR ABSENTEEISM DATA

Genetic algorithm for quantitative variables		Association rules phase		
		Association rules with agent	Association rules without agent	Number of rules extracted
Absenteeism data	Cluster 1	Time : 0.02 seconds. Memory : 38.16 MB.	Time : 0.30 seconds. Memory : 13.44 MB.	55
	Cluster2	Time : 0.01 seconds. Memory : 36.72 MB.	Time : 0.02 seconds. Memory : 16.63 MB.	10
	Cluster 3	Time : 0.01 seconds. Memory : 35.86 MB.	Time : 0.02 seconds. Memory : 13.11 MB.	3

Table VII presents a relevant improvement of the global execution time, including clustering and association rules. Concretely, as an average the execution time with the multi-agent integration is 8.33 times less than the traditional approach.

According to the experiments conducted in this section. The integration of multi-agent technology demonstrates the potential of sustaining the process of knowledge extraction from datasets. The results illustrated in Table VI confirm the efficiency of the developed multi-agent framework. However, the consumption of memory used requires an improvement.

The satisfying results obtained in this section with the three varied datasets, make our tool appropriate to solve the issue of agriculture domain covered by this research.

TABLE VII.   GLOBAL EXECUTION TIME

Data sets	Total execution time with agent	Total execution time without agent
Bio-degradation	0.08 seconds	0.84 seconds
Air quality	0.09 seconds	0.74 seconds
Absenteeism	0.10 seconds	0.73 seconds

IX.   APPLICATION OF THE FRAMEWORK FOR OLIVE CROPS DATASETS

The purpose of this section is mainly the evaluation of the rules returned by the autonomous mechanism of our multi-agent tool. Table VIII shows the eleven variables from the agricultural field in Morocco. Subsections of this part illustrate the results of both clustering and association rules agents.

TABLE VIII.   OLIVE OIL VARIABLES

Variables [min, max]		
Fertilization [0, 815]	Picking [1, 998]	Total cost [1.031, 997]
Phyotsanitary treatment [0, 624]	Irrigation [0, 885]	Depreciation cost [0, 769]
Ground work [0, 631]	Direct cost [1.236, 20227]	Olive total cost [0.14, 1.04]
Size [1.25, 625]	Indirect cost [0, 587]	
Total cost/ha [1.729,5072]	Production [730, 12187]	Yield [11.1, 26]
Transformation cost [0.025, 0.04]	Cost of transport [0.004, 0.1]	Olive oil total cost [1.26, 5.75] (variable to predict)

A. Result of Clustering Agent

Table IX shows three clusters generated. The analysis of this result illustrates, an improvement in the memory consumption compared to the three test case datasets. In addition, in this experiment, the clustering agent performs it task with a satisfactory execution time.

B. Result of Association Rules Agent

Table X presents sixty-sixth rules generated. In this experiment particularly in cluster 1, we note that, the execution time of association rules agent is higher compared to that in all test datasets. However, in this phase, there is an improvement in the memory consumption.

C. Interpretation

Table XI, illustrates the rules that satisfy the criteria of minimal olive oil cost. Note that, the rules extracted present high confidence values, which gives more credibility to the results.

From analysing, the rules obtained. Several rules indicate the impact of direct costs on the olive oil cost such fertilization, irrigation, picking and groundwork. In order to improve the value of olive agricultural chain and obtain lower cost of olive oil, the decision makers should focus on the optimization of direct costs.

TABLE IX.   RESULT OF CLUSTERING AGENT FOR OLIVE OIL DATASETS

Clusters	Variables	Execution time	Memory consumption
Cluster1	Fertilization, Phytosanitary treatment, Groundwork, size, Irrigation, Indirect cost, Depreciation cost, olive total cost, Yield, Transformation Cost, Cost of transport.	$3.39*10^{-4}$ seconds	14.72 MB
Cluster 2	Picking, direct cost, total cost, Total cost/ha,		
Cluster 3	Production		

TABLE X.   ASSOCIATION RULES RESULT FOR OLIVE OIL DATASETS

Clusters	Number of rules	Execution time	Memory consumption
Cluster 1	55	0.29 seconds	28.15 MB
Cluster 2	10	0.03 seconds	24.25 MB
Cluster 3	1	0.02 seconds	32.84 MB

TABLE XI.   EXTRACTED RULES

support = 20 (34%) , confidence = 90 % : Fertilization in [10.0; 123.0] --> Olive oil total cost in [1.85; 3.64]
support = 24 (41%) , confidence = 85 % : Size in [14.0; 145.0] --> Olive oil total cost in [1.54; 3.19]
support = 15 (25%) , confidence = 100 % : Ground work in [135.0; 244.0] --> Olive oil total cost in [1.97; 3.74]
support = 11 (18%) , confidence = 78 % : Picking in [336.0; 584.0] --> Transformation cost in [0.03; 0.083] AND Olive oil total cost in [1.94; 3.56]
support = 8 (13%) , confidence = 88 % : Fertilization in [10.0; 50.0] --> Transformation cost in [0.03; 0.083] AND Olive oil total cost in [1.85; 2.71]
support = 11 (18%) , confidence = 91 % Ground work in [10.0; 60.0] --> Transformation cost in [0.036; 0.083] AND Olive oil total cost in [1.6; 2.54]
support = 6 (10%) , confidence = 100 % : irrigation in [183.0; 264.0] --> Transformation cost in [0.025; 0.04] AND Olive oil total cost in [1.85; 2.67]

XI. CONCLUSION

In addition to the task of analyzing the olive oil, dataset and giving answers about parameters that optimize the cost of olive oil. The key goal of the Moroccan Green Plan is to increase agricultural production and farm income. In order to achieve this goal, public actions should optimize the organization of agricultural value chains and turn to promising sectors such the olive crops.

In the same context, this paper includes results about the factors that improve the value chain of olive oil production. Therefore, in this article, we propose an agent-mining framework that combined two data mining techniques in multi-agent environment. Thus, the use of the proposed framework can be extended into different domain that requires analyzing quantitative datasets.

The multi-agent framework includes data mining algorithms, Weka-based and JADE frameworks and four different types of agents precisely data agent, clustering agent, association rule agent and coordinator agent. The framework was defined to apply different data mining techniques using collaborative approach of interaction among agents to work with an integrated, intelligent perspective that primarily intend to improve the knowledge discovery process.

We tested our framework as potential solution to the issue of optimizing olive oil cost through three test cases from different domains. The results proved that the framework performed well with respect to runtime execution. Thus, the agent-mining tool proves it autonomous aspect and allows extracting sixty-sixth credible rules from olive oil datasets with performed execution time.

In future work, we will study the integration of real time datasets in our multi-agent tool with maintaining reduced runtime processing.

In the same agricultural context, the framework will be adapted to include instantaneous weather parameters and ground composition. Therewith, any upcoming change in the quality of minerals of the soil can be detected by the framework, which could send messages to the biological researchers in order to support preventive actions.

REFERENCES

[1] ADA, Agricultural Development Agency., "Les Fondements de la stratégie du Plan Maroc Vert." 2014.

[2] Faysse, N., and C. Simon. "Holding All the Cards Quality Management by Cooperatives in a Moroccan Dairy Value Chain."European Journal of Development Research, 2014.

[3] Obidallah, W.J., Raahemi, B. & Ruhi, U, "Clustering and Association Rules for Web Service Discovery and Recommendation: A Systematic Literature Review". SN COMPUT. SCI. 1, 27, 2020.

[4] University Waikato. Weka machine learning project, September 2009.

[5] R. Agrawal, T. Imielinski, and A. N. Swami, "Mining association rules between sets of items in large databases". In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 1993.

[6] F. Moslehi and A. Haeri, "A genetic algorithm-based framework for mining quantitative association rules without specifying minimum support and minimum confidence", Scientia Iranica D 27(3), pp.1316-1332, 2020.

[7] Gonzales, E., Mabu, S., Taboada, K., Shimada, K., Hirasawa, K., "Mining Multi-class Datasets using Genetic Relation Algorithm for Rule Reduction". In IEEE Congress on Evolutionary Computation, CEC 2009, pp. 3249–3255, 2009.

[8] Shah, S., Rashid, M. and Arif, M, "Estimating WCET using prediction models to compute fitness function of a genetic algorithm." *Real-Time* Syst 56, pp.28–63, 2020.

[9] Zayrit Soumaya, Belhoussine Drissi Taoufiq, Nsiri Benayad, Korkmaz Yunus, Ammoumou Abdelkrim,,"The detection of Parkinson disease using the genetic algorithm and SVM classifier", Applied Acoustics,Volume 171, 2021.

[10] Mesbahi, N., Kazar, O., Benharzallah, S., Zoubeidi, M.,"A Cooperative Multi-Agent Approach-Based Clustering in Enterprise Resource Planning". International Journal of Knowledge and Systems Science, 6(1) pp. 34-45, 2015.

[11] Gelvez Garcia, Nancy Yaneth; Balenduarte, Andrés David, Espitia Cuchango and Helbert Eduardo." Multi-Agent System Used for Recommendation of Historical and Cultural Memories". Tecciencia, vol.14, n.26, 2019.

[12] Calegari, R., Ciatto, G., Mascardi, V. *et al.*,"Logic-based technologies for multi-agent systems: a systematic literature review". Auton Agent Multi-Agent Syst 35, 1, 2021.

[13] Qian A., Chen CY.," Accurate Model of the Internet Financial Poverty Alleviation Based on the Multi-agent and the Data Mining". In Data Processing Techniques and Applications for Cyber-Physical Systems, DPTA 2019. Advances in Intelligent Systems and Computing, vol 1088. Springer, 2019.

[14] Popa, H.,Pop, D., Negru,V., Zahariae,D. ," Agent Discover : A Multi Agent System for knowledge discovery". In IEE/WIC/ACM International conference on Web Intelligence and Intelligent Agent Technology, Sydney,pp. 571-574 , 2008.

[15] Tong, C., Sharma, D., Shadabi, F." A Multi-Agents Approach to Knowledge Discovery". In IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Sydney, pp. 571–574, 2008.

[16] Nahar, J., Imam, T., Tickle, K., Chen, Y. "Association rule mining to detect factors which contribute to heart disease in males and females.", Expert Systems with Applications 40(4), pp. 1086–1093, 2013.

[17] Ait-Mlouk A., Agouti T., Gharnati F."Multi-agent-based modeling for extracting relevant association rules using a multi-criteria analysis approach", Vietnam J Comput Sci 3:235–245, 2016.

[18] Kaur, H., Chauhan, R., Alam, M. A,.Aljunid, S., Salleh, M.: SpaGRID, "A special Grid Framework for High Dimentional Medical Databases". In HAIS 2012, Part 1. LNCS, vol. 7208, pp.690-704, Springer, Heidelberg, 2012.

[19] F. Bellifemine, A. Poggi, and G. Rimassi. JADE,"A FIPA-Compliant agent framework", Proceedings Practical Applications of Intelligent Agents and Multi-Agents, 1999.

[20] Dua, D. and Graff, C., "UCI Machine Learning Repository", CA: University of California, School of Information and Computer Science, Irvine, 2019.

[21] A.Salleb-Aoussi, C.Vrain and C. Nortet, "Quantminer, a Genetic Algorithm for Mining Quantitative Association Rules ", IJCAI-07 1036, 2013.

[22] I.Belabed, M.Talibi Alaoui and A.Belabed, "Association Rules Algorithms for Data Mining Process Based on Multi Agent System". MLN 2019, LNCS 12081, pp. 431–443, 2020.