Expert Systems with Applications 39 (2012) 11243–11249

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# Web usage mining to improve the design of an e-commerce website: OrOliveSur.com

C.J. Carmona [a,*], S. Ramírez-Gallego [a], F. Torres [b], E. Bernal [c], M.J. del Jesus [a], S. García [a]

[a] Department of Computer Science, University of Jaén, 23071 Jaén, Spain
[b] Department of Marketing, University of Jaén, 23071 Jaén, Spain
[c] Department of Economics, University of Jaén, 23071 Jaén, Spain

ARTICLE INFO

*Keywords:*
Web usage mining
OrOliveSur.com
Subgroup discovery
Association rules
Clustering

ABSTRACT

Web usage mining is the process of extracting useful information from users history databases associated to an e-commerce website. The extraction is usually performed by data mining techniques applied on server log data or data obtained from specific tools such as *Google Analytics*. This paper presents the methodology used in an e-commerce website of extra virgin olive oil sale called www.OrOliveSur.com. We will describe the set of phases carried out including data collection, data preprocessing, extraction and analysis of knowledge. The knowledge is extracted using unsupervised and supervised data mining algorithms through descriptive tasks such as clustering, association and subgroup discovery; applying classical and recent approaches. The results obtained will be discussed especially for the interests of the designer team of the website, providing some guidelines for improving its usability and user satisfaction.



## 1. Introduction

Electronic commerce is the buying and selling of products or services through electronic media, such as Internet and other computer networks. Originally, the term was applied to the execution of transactions through electronic transactions such as electronic data interchange. However, with the advent of Internet in the mid 90's, it began mainly referring to the sale of goods and services on Internet, primarily using electronic payment. The amount of trade conducted electronically has grown extraordinarily since the spread of Internet. A high variety of commerce is made in this way (Soares, Peng, Meng, Washio, & Zhou, 2008), stimulating the creation and use of innovations such as electronic funds transfer, the supply chain management, marketing on Internet, online transaction processing, electronic exchange data, systems, inventory management and automated data collection.

After the concentration of olive oil cooperatives in Andalusia (Spain), in the last years, the literature proliferates on the export of olive products (Moral-Pajares & Lanzas-Molina, 2009), the use of e-commerce in the agricultural cooperatives and the adoption of Information and Communication Technologies as an essential tool in such export. Interesting studies on the international market and demand for olive oil were conducted by Mili and Zuniga (2003). Also, it is worth mentioning the discussion of factors affecting consumer demand for olive oil by selection bias models based on the Heckman correction (Tsakiridou, Mattas, & Tzimitra-Kalogianni, 2006). In addition, other contributions concerning studies located outside Spain, advocates of consumer satisfaction and costs of olive oil (Krystallis, Fotopoulos, & Zotos, 2006). Blery and Kapsopoulou (2007) conducted a study related to the promotion and marketing of a Greek company, olive oil and predict that exports are needed to increase sales.

The need arises to propose methodologies for intelligent data analysis, to enable the extraction of useful knowledge from the data (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). This is the concept of the Knowledge Discovery in Databases (KDD), which can be defined as the nontrivial process of identifying patterns in data with the following characteristics: valid, novel, useful and understandable (Han, 2005). The KDD process is a set of interactive and iterative steps, including among them the pre-processing of the data to correct inaccuracies, incompleteness or inconsistency present, reducing the number of records or finding the most representative features. KDD combines the traditional techniques of knowledge extraction with numerous resources developed in the area of artificial intelligence.

In the specialized literature, we found recent applications and consolidated reviews on the use of data mining in e-commerce. In Schafer, Konstan, and Riedl (2001), the authors discussed different models of e-commerce recommendation and in Hu and Liu (2004) a methodology to extract information from customer questionnaires was provided. The extraction of predictive knowledge is used to set personalized recommendations in web use (Zhang & Jiao, 2007) and association rules are used for descriptive same task (Lazcorreta, Botella, & Fernandez-Caballero, 2008). Predictive and descriptive tasks can hybridize to achieve the same purpose (Kim, Cho, Kim,

* Corresponding author. Tel.: +34 953 211956; fax: +34 953 212472.
*E-mail address:* ccarmona@ujaen.es (C.J. Carmona).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.03.046

Kim, & Suh, 2002) and the recommendation of time-varying products (Min & Han, 2005). The analysis of customers' behaviour in e-commerce websites has attracted much attention in recent years (Liao, Chen, & Lin, 2011; Thorleuchter, Poel, & Prinzie, 2012).

In this paper, we describe an specific methodology for extracting useful information from web usage data acquired using Google Analytics in the website www.OrOliveSur.com. This website is devoted to the national and international sales of extra virgin olive oil and derivatives products from the south region of Spain, Andalusia. The phases described are those associated to the extraction and preprocessing of the data, extraction of knowledge and analysis of results. Among the techniques employed in order to analyse this data set are used clustering, association rules and subgroup discovery techniques because we need to describe user behaviours in the website with respect to several properties. Some examples of association rules, clusters and interesting subgroups will be analyzed raising a discussion about the best ways for improving the usability and comfort for the customer of the website.

The paper is organised as follows: Section 2 presents the main information about the e-commerce website in which is based this paper "www.OrOliveSur.com", Section 3 shows a brief summary about web usage mining and advantages contributed by this technology, in Section 4 the complete experimental study is presented and finally, Section 5 presents concluding remarks about this study to the experts.

## 2. E-commerce website: OrOliveSur.com

OrOliveSur[1] is a project born in the province of Jaén from Andalusia (Spain) in 2010. The main purpose is to announce to the world the treasure of its land, the extra virgin olive oil. This website is focused in the olive oil produced in a particular territory of Jaén: the Sierra Mágina Natural Park. Sierra Mágina is a protected area of 50,000 acres of natural park, made up of forested slopes, concealed valleys and rugged mountain peaks. The highest peak, the Mágina Mountain is the highest in the Jaén province, standing at 2,167 metres.

OrOliveSur's catalog presents a wide number of extra virgin olive oils focused on the picual variety. This is the most extended olive grove variety at the world. In Spain it represents 50% of production. Most of it is to be found in Andalusia, especially in the province of Jaén. Its olive is large-sized and elongated in shape, with a peak at the end. The trees of this variety are of an intense silvery colour, open and structured. In addition, picual variety has excellent organoleptic properties because in stability and oleic acid obtains the best values with respect to other varieties like arbequina or hojiblanca, among others.

It is interesting to remark that users can find an English (http://en.orolivesur.com) and Spanish (http://www.orolivesur.com) version. In Fig. 1 the homepage of OrOliveSur is shown.

Along two years, OrOliveSur has received both national and international orders from European Union countries (Spain, Denmark, Germany, Great Britain, France, etc.), and its visits and orders are increased every day. The most important characteristic is that OrOliveSur offers a complete catalog with a lot of products and complete descriptions about these ones. For example, in Fig. 2 complete description for a product can be observed where technical data and several information of the olive oil are shown. Moreover, the OrOliveSur website gives direct sales and clients can pay by transfer bank, PayPal or credit card.

## 3. Web usage mining

Etzioni (1996) defined web mining as the use of data mining techniques to discover and extract knowledge in a website automatically, while Cooley, Mobasher, and Srivastava (1999) was further on highlighting the importance to consider the behaviour and preferences of the user. Anyway, different authors coincide to separate the web mining in several stages (Kosala & Bockeel, 2000; Liu, 2006):

1. Finding resources.
2. Selecting information and preprocessing.
3. Discovering knowledge.
4. Analysing patterns obtained.

Web mining can be classified in three domains with respect to the nature of data (Cooley, Mobasher, & Srivastava, 1997; Markov & Larose, 2007): web content mining, web structure data and web usage mining. This paper is focused in web usage mining which was defined by Srivastava, Cooley, Deshpande, and Tan (2000) as:

> The process of applying data mining techniques to the discovery of usage patterns from Web data.

Patterns are represented as a collection of pages or items visited by users. These patterns can be employed to understand the main features of the visitants behaviours in order to improve the structure of one website and create personal or dynamic recommendations about content of the web (Mobasher, 2005).

Web usage mining can be employed in several proposals like for example from to analyse pages sequences, quality of the website or search for global effectiveness. All proposals have been classified with respect to a taxonomy defined in Facca and Lanzi (2005) in four applications:

- Personalisation whose objective is based on recommendation systems, i.e. recommending interesting links to products which could be interesting to users.
- Pre-fetching and Caching which attempts to improve the performance of servers and applications loading possible pages in cache before than users request them.
- Design is related to the usability of a website. Studies in design can provide guidelines for improving the design of web applications and websites.
- E-commerce where techniques used within this group are related to Customer Relationship Management. The main purpose of these applications is to increase the sales of e-commerce websites.

## 4. Web usage mining in OrOliveSur.com

The main purpose of this experimental study is focused on the study of design in the e-commerce website of OrOliveSur.com through web usage mining techniques. These techniques are applied within KDD process which is divided in different phases. Specifically, in this experimental study we perform in a successive way the following phases:

- Compilation and pre-processing of data. This stage is presented in Section 4.1 and the main objective is to show how we have collected data and the pre-processing methods applied in order to prepare these ones for the next phase.
- Data mining. This phase is the most important in the process. To do so, we apply different descriptive data mining algorithms to the data which are presented in Section 4.2.
- Analysis and validation. This one validates and presents the results obtained for data mining algorithms used. In Section 4.3 we present the most important results obtained for different algorithms.

[1] http://www.orolivesur.com.


**Fig. 1.** Homepage from the e-commerce website OrOliveSur.com.

### 4.1. Compilation and pre-processing of data

Data set has been obtained with the webmaster tool *Google Analytics* from the period 1st January to 31st December for the year 2011. Moreover, several filters has been applied in data set in order to obtain only instances where bounce rate is lower than 100.00%. This value is the percentage of single-page visits or visits in which the person left your site from the landing page, i.e. we only obtain visits where users have been visited the website more than one seconds. In total the data set is composed by 8,832 instances.

In this experimental study are analysed several information related to the main features of visits. These properties are described below:

- Browser: This property contains the generic name of the browser used by the user in the visit. Among the possible values can be found the widely known Internet Explorer, Firefox, Chrome and Safari or mobile browser like Android or Blackberry®.
- Visitor type: It contains the type of the visitor. This property can contain the value of new (N) or returning (R) visitor.
- Keyword: It is the keyword access to the website by the user. The complete keyword set has been classified in six categories. It is important to remark that keywords of the original data set can be found in different languages and they are classified in a general category with English terms:
  - Olive oil. This value contains all the generic keywords related to the olive oil like *buy olive oil*, *venta de aceite*, *organic olive oil*, *aceite de oliva virgen extra*, *huile d'olive*, *Olivenöl Extra Vergine* and so on.
  - Iberian products. In this value are grouped all the generic keywords about iberian products like *jamón ibérico*, *buy iberico acorn-fed ham*, *jamón de bellota* and so on.
  - Brand. This keyword contains the entries related to any brand of the catalog as *La Casona*, *Verde Salud*, *Gamez-Piñar* or *OrOliveSur*, for example.


**Fig. 2.** Description of a product in OrOliveSur.com.

- Gift. It contains values related to gifts like *Wedding*, *Christmas Hamper* or *Cestas de Navidad*.
- Other. This value groups all access with keywords not classified within the previous ones.

- Nothing. Access without keyword are identified with this keyword as for example direct access.

- Source: This property indicates the source used by the user to access to the website.
  - Direct (D). This value is indicated for access performed directly to the website through the URL http://www.orolivesur.com.
  - Engine (E). Access performed through search engine like *Google*, *Yahoo* or *Live* have this value.
  - Mail (M). It indicates the access performed by an e-mail with a link to the website http://www.orolivesur.com.
  - Reference (R). This value can be found in access performed by other websites like directories, blogs and so on.
  - Social network (N). It contains all access performed through social network like *Facebook*, *YouTube*, *Twitter* or *Google Plus*.
- New visits: It indicates the number of new visits performed with the same browser, visitor type, keyword and source.
- Page views: It indicates the page views for users with the same browser, visitor type, keyword and source.
- Time on site: This feature indicates the time spent on website by users with the same browser, visitor type, keyword and source.
- Visits: This one shows the number of visits performed with the same browser, visitor type, keyword and source.
- Unique page views: It presents the number of unique page views by users with the same browser, visitor type, keyword and source.
- Page views per visit: This property indicates the complete number of page views for every visit.
- Unique page views per visit: It shows the complete number of unique page views for every visit.
- Time per page view: It shows the time used by the user per page view.

### 4.2. Data mining

Once data have been prepared, we apply different descriptive data mining techniques with the main objective of discovering knowledge which could give to the webmaster team, information to improve the design of the web and increase the number of visits received, in conclusion to have more orders and clients in the future. Between techniques employed we use clustering because it gives different behaviour patterns of the user which could help to understand the use of different browser or access, in addition we apply association rule learning because experts need to analyse rules without a target variable defined. Finally, experts have determined the use of subgroup discovery techniques because they are interested in obtaining information related to target variable such as keyword, visitor type and source.

Below, a brief description about techniques and algorithms used are presented:

- Clustering is the descriptive task of assigning a set of objects into several groups. We perform this task through the algorithm *k*-Means (Lloyd, 1982). The main objective of this algorithm is to partition the data set into *k* clusters in which each instance belongs to the cluster with the nearest mean (MacQueen, 1967). *k*-Means algorithm starts from *k* central points chosen randomly and every instance is assigned to the closer central point. Next, it performs a reassignment of the central points. Finally, algorithm is deemed to have converged when the assignments no longer change.

- Association rule learning (Agrawal, Imieliski, & Swami, 1993, 1996) is a descriptive data mining technique which attempts to discover interesting relations between variables in large databases using unsupervised learning. This technique is tackled with Apriori (Agrawal et al., 1993). This algorithm obtains association rule with the form $X \rightarrow Y$, where $X$ and $Y$ are itemsets. The set of all items in the domain of investigation, consisting of a set of transactions. Apriori considers that transactions correspond to training examples of a data set, an item is a binary feature, and itemsets are conjunctions of features. The quality of an association rule obtained by Apriori are defined by its confidence and support, i.e. Confidence of a rule is the conditional probability of $Y$ given $X$: $p(Y|X)$, and Support of a rule is an estimate of the probability of itemset $X \cup Y$: $p(X \cdot Y)$.
- Subgroup discovery task (Herrera, Carmona, González, & Del Jesus, 2011) is a data mining technique whose main purpose is to extract descriptive knowledge concerning a property of interest from the data (Kloesgen, 1996; Wrobel, 1997). In addition, this technique have been applied in different domains obtaining good results (Carmona, González, Del Jesus, Romero, & Ventura, 2010, 2011; Carmona, González, Del Jesus, Navío, & Jiménez, 2011; Romero, González, Ventura, Del Jesus, & Herrera, 2009). The algorithm used within this technique is the most representative throughout the literature, the NMEEF-SD algorithm (Carmona, González, Del Jesus, & Herrera, 2010). This algorithm is an evolutionary fuzzy system whose objective is to extract descriptive fuzzy and/or crisp rules for the SD task, depending on the type of variables present in the problem. It is based on NSGA-II approach (Deb, Pratap, Agrawal, & Meyarivan, 2002) and it can use a wide number of quality measures for subgroup discovery as objectives.

In Table 1 are described parameters used by algorithms described previously.

### 4.3. Analysis and validation

In this section are presented results obtained by different algorithms studied with the data set `OrOliveSur.com`. In order to understand the results obtained in the best way its results and analysis are shown in separate sections for each algorithm.

#### 4.3.1. Clustering through k-Means

*K*-Means algorithm obtains the number of clusters indicated by users. Specifically, in this experimental study we obtain 5 clusters where for each one different properties are considered. Table 2 shows properties and values of the clusters.

As can be observed in Table 2 main clusters are obtained with respect to a common property, the use of search engine to access to the website because all clusters contain this value. With these results is interesting to highlight the following statements:

- Cluster A (11% instances) groups all access performed by Internet Explorer through a search engine where keywords were not obtained. In this way, webmasters must work to obtain these keywords because these users visit the website during a wide period of time and they are possible potential customers.
- Cluster with the major values for time and page views is the Cluster B (10% instances). It is interesting to remark that in this cluster, users employ Firefox and the keyword Brand in order to access. Due to the visitor type are returning is possible that the main access to the website is through the keyword "OrOliveSur" because they know this e-commerce website. Cluster E (17%) has similar properties with respect to Cluster B but in this case with direct and search engine access.
- Clusters obtained by *k*-Means with the major number of instances grouped are: Cluster C with 34% of instances and Cluster D with 27%. It is interesting to highlight that the main property of these clusters is the access through keyword olive oil. In this way, we recommend to webmasters to analyse the importance of olive oil versus iberian products in the website because the majority access are performed by olive oil and it is probably that they need to improve the position of keywords related to iberian products.

#### 4.3.2. Association rule learning with Apriori algorithm

Apriori algorithm obtains a huge number of rules with high quality, i.e. with high values of confidence, but the majority of these rules show that algorithm functions correctly and does not show an interesting behaviour in the access. Rules more representative obtain by Apriori are show in Table 3.

All rules shown in Table 3 have the highest value of confidence with 100%, i.e. all examples covered by rules are correctly covered. In addition, support obtained by these rules have interesting values because they are upper than 10%. Among interesting rules obtained by Apriori algorithm we can analyse them in three groups:

- This group is formed by rules $R1$ and $R2$ where accesses performed through search engine are presented. First rule ($R1$) presents a high access of users with browser Internet Explorer and with keyword olive oil and the second one ($R2$) shows a high level of access from search engine without keyword identified. In this way, all the statements obtained by *k*-Means algorithm are confirmed with this group of rules obtained by Apriori.
- Rules $R3$, $R4$, and $R5$ present the access to the website with keyword olive oil. These rules have high values of support. In this way, we complete the previous analysis performed and we recommend to the experts reviews the position in iberian products of OrOliveSur.com because the majority access are to buy olive oil.
- Finally, this group is based in rule $R6$. This rule indicates with a 100% of confidence that users access directly to the website are returning visitor type. It is really interesting because the 10% of support indicates that this percentage of users knew OrOliveSur.com previously. With these results, we would recommend to analyse the website and to be careful with the changes in the design because these ones could confuse to habitual clients.

**Table 1**
Parameters of algorithms employed.

| Algorithm | Parameters |
|---|---|
| *k*-Means | $k = 5$ |
| Apriori | Minimum support = 0.1, minimum confidence = 0.9, bins = 10 |
| NMEEF-SD | Population size = 50, evaluations = 10000, crossover probability = 0.60, mutation probability = 0.1, minimum confidence = 0.6, representation of the rule = canonical, linguistic labels = 9, objective1 = sensitivity and objective2 = unusualness. |

**Table 2**
Properties of clusters obtained by $k$-Means ($k = 5$).

| Attribute | Cluster A | Cluster B | Cluster C | Cluster D | Cluster E |
|---|---|---|---|---|---|
| Browser | IE | FI | CH | IE | FI |
| Visitor type | N | R | N | N | R |
| Keyword | Nothing | Brand | Olive oil | Olive oil | Nothing and brand |
| Source | E | E | E | E | D and E |
| Page views | 7.26 | 10.02 | 6.66 | [5.54,9.61] | 5.68 |
| Time | 463.43 | 822.09 | 414.04 | [266.30,623.58] | 402.69 |
| Unique page views | 4.65 | 5.65 | 4.33 | [3.79,6.03] | 3.75 |

**Table 3**
Rules obtained by Apriori algorithm.

| ♯ | Rule | Support | Confidence |
|---|---|---|---|
| *R*1 | IF browser = IE AND visitor type = N AND keyword = olive oil THEN source = E | 0.112 | 1.000 |
| *R*2 | IF keyword = nothing AND unique page views = (1.5—2.5] AND unique page/visits = (1.92-2.07] THEN source = E | 0.108 | 0.999 |
| *R*3 | IF keyword = olive oil THEN visits = (-*inf*-1.5] | 0.284 | 0.999 |
| *R*4 | IF browser = IE AND keyword = olive oil THEN visits = (-*inf*-1.5] | 0.160 | 0.999 |
| *R*5 | IF visitor type = N AND keyword = olive oil THEN visits = (-*inf*-1.5] | 0.194 | 0.999 |
| *R*6 | IF source = D THEN visitor type = R | 0.101 | 1.000 |

**Table 4**
Rules and results obtained by NMEEF-SD algorithm.

| ♯ | Rule | SIGN | UNUS | SENS | FCNF |
|---|---|---|---|---|---|
| *R*1 | IF source = E THEN keyword = olive oil | 1949.707 | 0.117 | 0.999 | 0.483 |
| *R*2 | IF source = E THEN keyword = brand | 1949.707 | 0.073 | 1.000 | 0.303 |
| *R*3 | IF time/page views = Low THEN keyword = nothing | 3.920 | 0.001 | 0.999 | 0.448 |
| *R*4 | IF time = Low THEN keyword = nothing | 11.175 | 0.005 | 0,982 | 0.486 |
| *R*5 | IF keyword = nothing AND page views = Very low AND unique page views = Very low THEN source = R | 2216.810 | 0.090 | 0.996 | 0.373 |
| *R*6 | IF keyword = nothing AND unique page views = Very low THEN source = R | 2265.863 | 0.089 | 0.999 | 0.368 |
| *R*7 | IF keyword = nothing AND page views = Very low AND page/visits = Very low THEN source = R | 2216.810 | 0.090 | 0.996 | 0.372 |
| *R*8 | IF keyword = nothing AND unique page views = Very low AND unique page/visits = Very low THEN source = R | 2265.863 | 0.089 | 0.999 | 0.368 |
| *R*9 | IF visitor-type = N AND unique page views = Low THEN source = E | 90.077 | 0.038 | 0.658 | 0.653 |
| *R*10 | IF browser = IE AND page views = Low THEN source = E | 137.419 | 0.057 | 0.575 | 0.709 |
| *R*11 | IF new visits = 0 THEN visitor type = R | 2819.825 | 0.229 | 1.000 | 1.000 |

#### 4.3.3. Subgroup discovery through NMEEF-SD algorithm

The main purpose of subgroup discovery is to find interesting relationships in data with respect to an interest property. Specifically for this problem, we analyse properties as keyword, source of visitor type, for example.

Below, the most relevant subgroups obtained for NMEEF-SD algorithm with respect to the different property values together values of quality measures are shown in Table 4. This one describes rules obtained and the quality measures of significance (*SIGN*), unusualness (*UNUS*), sensitivity (*SENS*) and fuzzy confidence (*FCNF*). A complete description of these quality measures can be found in Herrera et al. (2011).

As can be observed in results obtained by NMEEF-SD, there are a huge number of rules with high values in the majority of quality measures. Even though some rules like *R*11 is obvious because if visits are not news the consequence is because users are returning. However, this rule provides information about the good behaviour of the algorithm used.

It is interesting to remark that users what access directly to the website, i.e. without using any keywords as rules *R*3 and *R*4 show in the results, remain in the website during an acceptable time and time per page views is interesting. In addition, *R*5, *R*6, *R*7 and *R*8 show that reference websites like directories or blogs with external links to OrOliveSur are visits with low number of page-views and unique-page-views. In this way, webmaster must improve the description and image of OrOliveSur in these reference websites because it is probably that users does not find the information hoped.

Rule most interesting discovered by NMEEF-SD is the use of the browser Internet Explorer for the majority users that visit OrOliveSur through search engine as *Google* or *Yahoo*, for example. These users visit between 1 and 100 pages in the website. In this way, we recommend to the webmaster to analyse the design of the website to test that is correctly shown and designed in this browser in different versions.

## 5. Concluding remarks

In this paper, a study based on data mining techniques in order to extract knowledge in a data set with information about users history associated to an e-commerce website is presented. These data are collected from the e-commerce website `OrOliveSur.com` which is related to the sell of extra virgin olive oil and iberian products from Spain. The main purpose is to apply a set of descriptive data mining techniques to obtain rules that allow to help to the webmaster team to improve the design of the website. Techniques used are clustering, association rules and subgroup discovery. Therefore, in this web usage mining study can be analysed different knowledge with respect to these techniques used, i.e. we present a variety of groups and/or rules associated for each technique.

In general, knowledge discovered is related to the original point of access of the users. Concretely, clustering technique k-Means obtains interesting knowledge for webmaster team because there are a wide number of access performed with Internet Explorer through search engine without keywords. In this way, webmaster team must study this problem because more than 10% of access

are generated in this way. With respect to the rules and subgroups obtained through Apriori and NMEEF-SD different conclusions are obtained: first, webmaster team must pay attention in visits obtained by reference websites because users visit a very low number of pages and second the majority of visits use Internet Explorer. In this way webmaster team must improve the image of `OrOliveSur.com` in this type of websites and browser.

Finally, results obtained show that webmaster team must earnestly improve keywords related to iberian products because there is not information about the access of users through these keywords.

## Acknowledgment

This paper was supported by the Spanish Ministry of Education, Social Policy and Sports under project TIN-2008-06681-C06-02, FEDER Founds, by the Andalusian Research Plan under project TIC-3928, FEDER Founds, and by the University of Jaén Research Plan under project UJA2010/13/07 and Caja Rural sponsorship.

## References

Agrawal, R., Imieliski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on management of data* (pp. 207–216). ACM Press.

Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. (1996). Fast discovery of association rules. In U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances in knowledge discovery and data mining* (pp. 307–328). AAAI Press.

Blery, E., & Kapsopoulou, K. (2007). Marketing olive oil: A case study from Greece. *Journal of Food Products Marketing, 13*, 39–55.

Carmona, C. J., González, P., Del Jesus, M. J., & Herrera, F. (2010). NMEEF-SD: Non-dominated multi-objective evolutionary algorithm for extracting fuzzy rules in subgroup discovery. *IEEE Transactions on Fuzzy Systems, 18*, 958–970.

Carmona, C. J., González, P., Del Jesus, M. J., Navío, M., & Jiménez, L. (2011). Evolutionary fuzzy rule extraction for subgroup discovery in a psychiatric emergency department. *Soft Computing, 15*, 2435–2448.

Carmona, C. J., González, P., Del Jesus, M. J., Romero, C., & Ventura, S. (2010). Evolutionary algorithms for subgroup discovery applied to e-learning data. In *Proceedings of the IEEE international education engineering* (pp. 983–990).

Carmona, C. J., González, P., Del Jesus, M. J., & Ventura, S. (2011). Subgroup discovery in an e-learning usage study based on Moodle, In *Proceedings of the international conference of European transnational education* (pp. 446–451).

Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the World Wide Web. *On Tools with Artificial Intelligence*, 558–567.

Cooley, R., Mobasher, B., & Srivastava, J. (1999). Data preparation for mining World Wide Web browsing patterns. *Knowledge and Information Systems, 1*, 5–32.

Deb, K., Pratap, A., Agrawal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions Evolutionary Computation, 6*, 182–197.

Etzioni, O. (1996). The World Wide Web: Quagmine or gold mine. *Communications of the ACM, 39*, 65–68.

Facca, F. M., & Lanzi, P. L. (2005). *Mining Interesting Knowledge from Weblogs: A Survey, 53*, 225–241.

Fayyad, U. M., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery: An overview. In *Advances in knowledge discovery and data mining* (pp. 1–34). AAAI/MIT Press.

Han, J. (2005). *Data mining: Concepts and techniques*. Morgan Kaufmann Publishers Inc.

Herrera, F., Carmona, C. J., González, P., & Del Jesus, M. J. (2011). An overview on subgroup discovery: Foundations and applications. *Knowledge and Information Systems, 29*, 495–525.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 168–177).

Kim, J. K., Cho, Y. H., Kim, W. J., Kim, J. R., & Suh, J. H. (2002). A personalized recommendation procedure for Internet shopping support. *Electronic Commerce Research and Applications, 1*, 301–313.

Kloesgen, W. (1996). Explora: A multipattern and multistrategy discovery assistant. In *Advances in knowledge discovery and data mining* (pp. 249–271). American Association for Artificial Intelligence.

Kosala, R., & Bockeel, H. (2000). Web mining research: A survey. *SIGKDD Explorations, 2*, 1–15.

Krystallis, A., Fotopoulos, C., & Zotos, Y. (2006). Organic consumers' profile and their willingness to pay (WTP) for selected organic food products in Greece. *Journal of International Consumer Marketing, 19*, 81–106.

Lazcorreta, E., Botella, F., & Fernandez-Caballero, A. (2008). Towards personalized recommendation by two-step modified Apriori data mining algorithm. *Expert Systems with Applications, 35*, 1422–1429.

Liao, S. H., Chen, Y. J., & Lin, Y. T. (2011). Mining customer knowledge to implement online shopping and home delivery for hypermarkets. *Expert Systems with Applications, 38*, 3982–3991.

Liu, B. (2006). *Web data mining: Exploring hyperlinks, contents, and usage data (data-centric systems and applications)*. Springer-Verlag.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory, 28*, 129–137.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley symposium on mathematical statistics and probability* (pp. 281–297).

Markov, Z., & Larose, D. T. (2007). *Data mining the web. Uncovering patterns in web content, structure and usage*. Wiley-Interscience.

Mili, S., & Zuniga, M. R. (2003). The olive oil sector facing new international market challenges: A demand-driven perspective. *Journal of International Food and Agribusiness Marketing, 14*, 35–55.

Min, D. H., & Han, I. (2005). Detection of the customer time-variant pattern for improving recommender systems. *Expert Systems with Applications, 28*, 189–199.

Mobasher, B. (2005). *Web usage mining and personalization*. CRC Press, LLC.

Moral-Pajares, E., & Lanzas-Molina, J. R. (2009). La exportacion de aceite de oliva virgen en Andalucia: Dinamica y factores determinantes. *Revista de Estudios Regionales*, 86.

Romero, C., González, P., Ventura, S., Del Jesus, M. J., & Herrera, F. (2009). Evolutionary algorithm for subgroup discovery in e-learning: A practical application using Moodle data. *Expert Systems with Applications, 36*, 1632–1644.

Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). E-commerce recommendation applications. *Data Mining and Knowledge Discovery, 5*, 115–153.

Soares, C., Peng, Y., Meng, J., Washio, T., & Zhou, Z. H. (Eds.). (2008). *Applications of data mining in e-business and finance. Frontiers in artificial intelligence and applications*. IOS Press.

Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. (2000). Web usage mining: Discovery and applications of usage patterns from web data. *SIGKDD Explorations*, 12–23.

Thorleuchter, D., Poel, D. V. D., & Prinzie, A. (2012). Analyzing existing customers' websites to improve the customer acquisition process as well as the profitability prediction in B-to-B marketing. *Expert Systems with Applications, 39*, 2597–2605.

Tsakiridou, E., Mattas, K., & Tzimitra-Kalogianni, I. (2006). The influence of consumer characteristics and attitudes on the demand for organic olive oil. *Journal of International Food and Agribusiness Marketing, 18*, 23–31.

Wrobel, S. (1997). An algorithm for multi-relational discovery of subgroups. In *Proceedings of the 1st European symposium on principles of data mining and knowledge discovery* (pp. 78–87). Springer.

Zhang, Y., & Jiao, J. (2007). An associative classification-based recommendation system for personalization in b2c e-commerce applications. *Expert Systems with Applications, 33*, 357–367.