Information mining from health protection data against mosquitoes

Stavros Valsamidis[1], Lambros Tsourgiannis[2], Giannoula Florou[1], Athanasios Mandilas[1], Michael Nikolaidis[1]

[1]Department of Accounting and Finance, EMaTTech, Kavala, Greece
e-mail: svalsam@teikav.edu.gr, ltsourgiannis@gmail.com, gflorou@teikav.edu.gr, smand@teiemt.gr, mnikol@teikav.edu.gr
[2]Directorate of Public Health and Social Care of Regional District of Xanthi, Region of Eastern Macedonia and Thrace, Xanthi, Greece

Abstract. West Nile Virus (WNV) first time appeared in 2012 in the Regional District of Xanthi in Greece in Northern Eastern Greece and in total 63 cases were reported between 2012-2014. The Region of Eastern Macedonia and Thrace (EMTh) conducted a mosquito and vector control program. In this study, we apply three different data mining techniques to data about 314 households and concerned 2014. They were analyzed using the classification algorithm PART, the clustering algorithm k-means and the association rule mining algorithm, Apriori from the WEKA data mining package. The results indicate that the infected persons have some common characteristics. Hopefully, this aims to generate detailed knowledge of household use of insecticide consumer products to kill mosquitoes, the factors that affect the use of these products. The findings indicate that there is room for improvement towards the self protection methods against mosquitoes.

Keywords: Mosquitoes, Protection, Classification, Clustering, Association Rule Mining.

1 Introduction

Mosquito borne diseases are major public health problems in many countries in all over the world. Regions in countries such as India, Mexico, Thailand, USA reported many different diseases caused by mosquito bites (Rosendaal, 1997; Raghavendra et al., 2011; Pandit et al., 2010). In recent years, Greece suffered by WNV which was responsible for many serious diseases and even deaths which reported during the period 2012-2015.

West Nile Virus (WNV) is a widespread disease mainly transmitted by mosquitoes. In particular, West Nile Virus (WNV) outbreak occurred in the Region of Eastern Macedonia and Thrace in 2012-2015. WNV infection can be asymptomatic or symptomatic in humans, with 4:1 ration (Center of Disease Control

Proceedings of the 8th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2017), Chania, Greece, 21-24 September, 2017.

and Prevention, 2015). This virus is transmitted by mosquitoes and can cause illness which can me mild resulting in influenza – like symptoms or severe affecting the central nervous system causing encephalitis (Lorono-Pino et al., 2014; Jones et al., 2014). In many WNV outbreaks reported deaths (Jones et al. 2014; He et al., 2014).

While humans are considered dead-end hosts once infected, birds have been documented to produce high enough levels of the virus to spread WNV to mosquitoes. Thus, bird populations produce a significant impact upon the growth of the disease. Approximately 80% of cases in humans show no noticeable symptoms, and the infected recover on their own. Another 20% develop mild symptoms similar to a flu. A serious neurological illness occurs in less than 1% of the infected population. Currently, there is no cure or preventative shot for this disease. Preventative measures, including killing off mosquitoes and minimizing personal exposure to mosquitoes, are the most effective ways to combat WNV.

WNV not only posses risk to health but diseases in endemic areas place a burden on households, on health services and the economic growth of the local communities (Koenraadt et al., 2006; Tyagi et al., 2005). Therefore, citizens' protection against mosquito bites is very important for public health.

Whilst several studies demonstrated the efficacy of mosquito management in response to WNV, Carney et al. (2005), Barber et. al. (2010) mainly suggested a reduction in human WNV cases associated with the application of mosquito control programs. Nowadays studies have revealed that citizens' knowledge, attitude, and practice of various methods of personal and household protection against mosquito bites vary in different endemic regions of tropical countries (Pandit et al., 2010). Various methods for protection from mosquito bites are used globally including repellent oils, smoldering coils, vaporizing mats, repellent creams, liquid vaporizer (Raghavendra et al., 2011). The market in these products worldwide is worth about 2 billion USD per year (WHO, 1998). Effectiveness of these methods lasts between five to seven hours with 60-80% protection (Curtis et. al., 1989; Ansari et al., 1990).

The motivation is the effective control of the infectious diseases transmitted by the inset particularly mosquito (Pandit et al., 2010). The use of personal or household protection methods are indicators of socioeconomic status, which has been reported as an important factor associated with diseases transmitted by mosquitoes and more particular with malaria (Tyagi et al., 2005). They also argued that the high usage of mosquito repellents by urban respondents and the low usage in rural respondents is explained the impact of socioeconomic conditions on the selection of protection means in communities (Tyagi et al., 2005). Moreover, education and knowledge of protection from mosquito bites, the promotion of health education and the positive role of women and family members in community interventions must be emphasized, is associated with less malaria infection (Tyagi et al., 2005; He et al., 2014). Another study aimed to identify the association between demographic characteristics (including age, sex, education, occupation, sub-district), knowledge of the population on symptoms of dengue, vector and prevention against mosquitoes; and practices such as container protection and mosquito reduction (Koenraadt et al., 2006).

The common household use of insecticide consumer products to kill mosquitoes: aerosol spray cans with insecticide were used to kill mosquitoes in 70% of homes, and insecticide emitters were used in 10–20% of homes (Lorono-Pinoet al., 2013). This heavy use of insecticide consumer products is not surprising in light of our previous reports of large numbers of Ae. aegypti and another human-biting mosquito, Culex quinquesfasciatus, being present in homes in Merida City (Garcia-Rejon et al., 2008; Lorono-Pino et al., 2013). Other studies have reported use of insecticide consumer products for 28–89% of households in dengue endemic settings in Asia (van Benthem et al., 2002; Itrat et al., 2008; Syed et al., 2010; Naing et al., 2011; Al-Dubai et al., 2013; Mayxay et al., 2001) or the Americas (Shuaib et al., 2010). This is unfortunate because, as shown by a recent study from a malaria-endemic area in Africa, much can be learned from in-depth assessments of household use of pest control products (Nalwanga and Ssempebwa, 2011). Moreover, there are potential negative health effects, particularly for asthma and respiratory diseases, from inhalation of pesticide aerosols or vapors (Hernandez et al., 2011). Improved knowledge of the extent of household use of insecticide consumer products is important not only to determine the willingness of households to invest in the use of domicile-targeted insecticide-based products – to kill mosquitoes, cockroaches, and other indoor pests – but also to help assess the overall insecticide exposure in the environment stemming from household use, vector control program applications to suppress mosquitoes or other arthropods spreading pathogens to humans or domestic animals, and agricultural applications to protect crops.

On the other hand, data mining is an iterative process of creating predictive and descriptive models, by uncovering previously unknown trends and patterns in vast amounts of data, in order to extract useful information and support decision making (Kantardzic, 2003). Data mining methods are divided into three major categories (Witten & Frank, 2005). The first category involves the classification methods, whereas the second the clustering ones and the third the association rule mining methods. Classification methods use a training dataset in order to estimate some parameters of a mathematical model that could in theory optimally assign each case from a new dataset into a specific class. In other words, the training set is used to train the classification technique how to perform its classification. Clustering refers to methods where a training set is not available. Thus, there is no previous knowledge about the data to assign them to specific groups. In this case, clustering techniques can be used to split a set of unknown cases into clusters. Association rule mining discovers relationships, sometimes hidden, among attributes (variables) in a dataset.

Data mining techniques have already been applied for analyzing swarms. Swarm Intelligence is quite an emerging area of research (Timmis et al., 2010). A swarm is a large number of homogenous, simple agents interacting locally among themselves, and their environment, with no central control to allow a global interesting behaviour to emerge. Example of swarm is considered the population of mosquitoes. An early warning system for West Nile virus (WNV) outbreaks provides a basis for targeted public education and surveillance activities as well as timelier larval and adult mosquito control (Mostashari et al., 2003). They adapted the spatial scan statistic for prospective detection of infectious disease outbreaks, applied the results to data on dead birds reported from New York City in 2000, and reviewed its utility in providing an early warning of WNV activity in 2001. Data mining techniques were applied for dengue infection in order to correctly classify the patients since these classes require different treatment (Thitiprayoonwongse et al., 2012). Dengue infection is an epidemic disease typically found in tropical region. Decision trees was the main technique for the prediction of day0 date which is the critical date of dengue patients that some patients face the fatal condition..

The present study was conducted to assess the awareness and practices of mosquito bite prevention methods among households of REMTH by using data mining techniques. Total 314 households have participated in the study from REMTH area. Telephonic interviews using a structured questionnaire performed to all households. The study was conducted in the month of August 2014. The pilot pre-tested structure questionnaire was used to collect the data. Study respondents were 57% male and 43% female. Almost 99% had knowledge about breeding places of mosquito, but poor knowledge about biting time (20%). 71% of participants knew that mosquito bite causes WNV. 39% of households were using mosquito net as protection against the bite, but only 10% were using insecticide treated bed net. There is need of increasing use of insecticide treated bed nets and continuous updating of knowledge about various aspects of mosquito bite.

Section 2 describes the dataset and the methodology used in our research. Sections 3 presents the results from the data mining methods and Section 4 discusses the results and refers to the main conclusions of our research.

2 Data and Methodology

In this section the dataset we used in our methodology is described in detail. Also, the data mining methods applied to the mosquitoes data are explained and analyzed.

2.1 The Dataset

The dataset was collected from the Directorate of Public Health and Social Welfare of Xanthi. The data were collected during 2014 and involve 340 households from REMTh. The data are originally in ASCII form. Each household is described by 7 variables. The seven variables which are used in the analysis described in the methodology section are Infected, Age, No_Children, Family_Members, Occupation, Education, Exp_Selfprot_Program. Table 1 describes each variable in detail.

Table 1. The variables used in our analysis

Variable Name	Description	Type
Infected	If there was an infected by WNV person (0: No, 1: Yes)	Nominal
Age	The age of the respondent	Numeric
No_Children	The number of children of the respondent	Numeric
Family_Members	The number of family members of the respondent	Numeric
Occupation	The occupation of the respondent	Nominal
Education	The education level of the respondent	Nominal
Exp_Selfprot_Program	The expenses of the respondent for self-protection measures	Numeric

2.2 Data mining techniques

The WEKA (*Waikato Environment for Knowledge Analysis*) (Witten & Frank, 2005) computer package was used in order to apply classification, clustering and association rule mining methods to the dataset. WEKA is open source software that provides a collection of machine learning and data mining algorithms. Fig. 1 shows the basic Graphical User Interface (GUI) of WEKA. One of the main objectives of WEKA is to mine information from existing agricultural datasets (Cunningham and Holmes, 1999) and the main reason for choosing

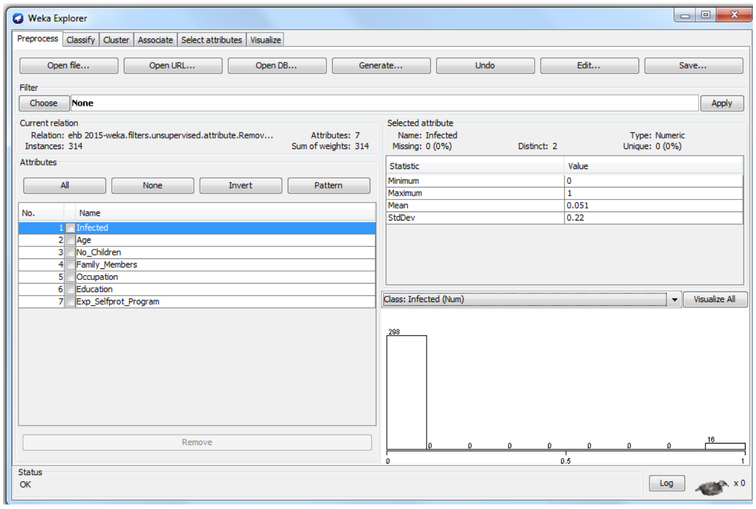

Fig. 1. WEKA environment

There are various classification methods implemented in WEKA, like ZeroR, OneR, PART etc. In the *classification* step, the algorithm *PART* (Witten & Frank, 2000) was applied to our data. It generates a decision list by using the separate-and-conquer method and builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule. PART can parsimoniously discover and represent simple relationships between the real data (Cunningham and Holmes, 1999). In our case the variable "Infected" is used as a class and shows whether a person was infected or not by WNV.

The *clustering* step uses the k-means algorithm (MacQueen, 1967; Kaufmann & Rousseeuw, 1990), called *SimpleKMeans* in WEKA. K-means is an efficient partitioning algorithm that decomposes the data set into a set of k disjoint clusters. It is a repetitive algorithm in which the items are moved among the various clusters until they reach the desired set of clusters. With this algorithm a great degree of similarity for the items of the same cluster and a large difference of items, which belong to different clusters, are achieved. The variable "Infected" is used in order to assess the accuracy of the clustering and investigate its impact on olive tree cultivation.

Association rule mining is one of the most well studied data mining tasks. It discovers relationships among attributes (variables) in datasets, producing if-then statements concerning attribute-values (Agrawal et al., 1993). An association rule X ⇒ Y expresses a close correlation among items in a dataset, in which transactions in the dataset where X occurs, there is a high probability of having Y as well. In an association rule X and Y are called respectively the antecedent and consequent of the rule. The strength of such a rule is measured by values of its support and confidence. The *confidence* of the rule is the percentage of transactions with antecedent X in the dataset that also contain the consequent Y. The *support* of the rule is the percentage of transactions in the dataset that contain both the antecedent and the consequent Y in all transactions in the dataset.

The WEKA system has several association rule-discovering algorithms available. The Apriori algorithm (Agrawal & Srikant, 1994) is used for finding association rules over the discretized LMS data table in Appendix 1. Apriori is the best-known algorithm to mine association rules. It uses a breadth-first search strategy to counting the support of item sets and uses a candidate generation function, which exploits the downward closure property of support. Iteratively reduces the minimum support until it finds the required number of rules with the given minimum confidence.

There are different techniques of categorization for association rule mining. Most of the subjective approaches involve user participation in order to express, in accordance with his/her previous knowledge, which rules are of interest. One technique is based on unexpectedness and actionability (Liu et al, 1996; Liu et al, 2000). *Unexpectedness* expresses which rules are interesting if they are unknown to the user or contradict the user's knowledge. *Actionability* expresses that rules are interesting if users can do something with them to their advantage. The number of rules can be decreased to unexpected and actionable rules only (García et al., 2007). Another technique proposes the division of the discovered rules into three categories (Minaei-Bidgoli et al., 2004). (1) *Expected and previously known*: This type of rule confirms user beliefs, and can be used to validate our approach. Though perhaps already known, many of these rules are still useful for the user as a form of empirical verification of expectations. (2) *Unexpected*: This type of rule contradicts user beliefs. This group of unanticipated correlations can supply interesting rules, yet their interestingness and possible actionability still requires further investigation. (3) *Unknown*: This type of rule does not clearly belong to any category, and should be categorized by domain specific experts.

3 Results

The first step before applying the data mining methods described in the previous section is the pre-processing of the data in order to prepare them for data analysis.

3.1 Pre-processing

The filter *NumericalToNominal* was applied to the data in order to convert numeric variables and their values to nominal. For example, number 0 and 1 in variable Infected are converted to nominal, where 0 signifies not infected and 1 infected. Fig. 2 depicts all the variables used in our analysis. The two colours correspond to Infected (Red) and not Infected (Blue).It is noteworthy that only 16 respondents were infected by WNV.

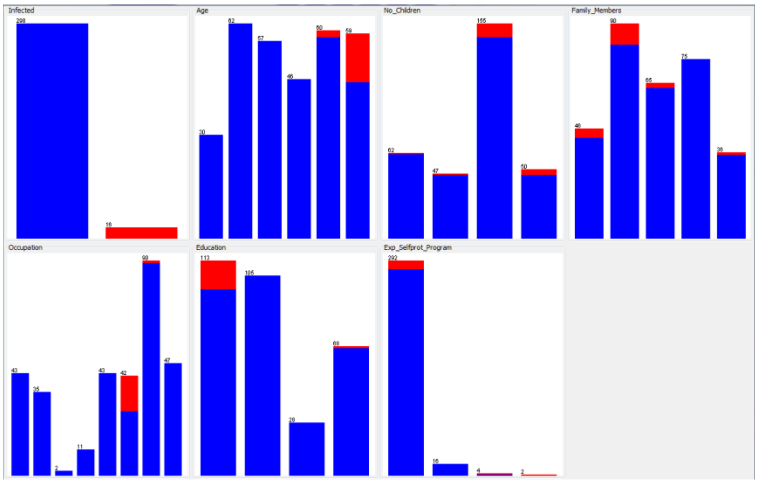

Fig. 2. Visualization of the attributes with class variable "Infected"

It is worth mentioning that all the sixteen infected persons belong to the same Age group, have the same Education level and the same Occupation but they have spent different amount of money for self protection. These findings are depicted on Fig3a, Fig3b, Fig3c and Fig3d.

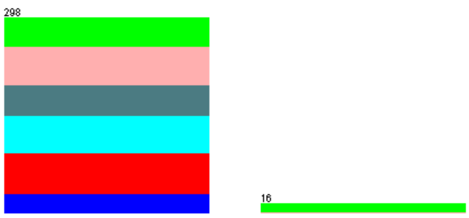

Fig. 3a. Visualization of the attribute Age with class variable "Infected" (the 16 infected persons correspond to the older people)

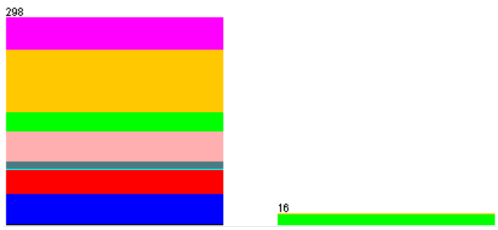

Fig. 3b. Visualization of the attribute Occupation with class variable "Infected" (the 16 infected persons correspond to the farmers)

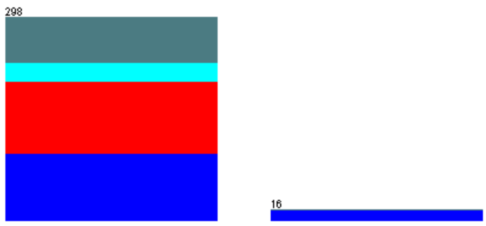

Fig. 3c. Visualization of the attribute Education with class variable "Infected" (the 16 infected persons correspond to the people of low education)

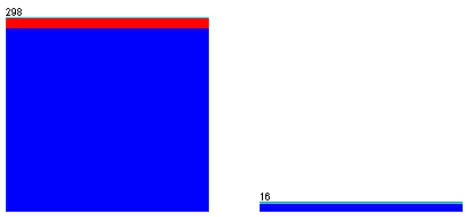

Fig. 3d. Visualization of the attribute Exp_SelfProt_Program with class variable "Infected" (most of the 16 infected persons correspond to the people who do not expend money for self-protection against mosquitoes)

3.2 Classification

In the classification step, the algorithm *PART* is applied which generates a decision list. It uses separate-and-conquer method and builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule. The attribute "Infected" is used as a class.

Table 3. Classification results using variable "Infected" as class.

PART decision list
Occupation = 7: 0 (90.29/1.0)
Age = 2: 0 (62.0)
Age = 3: 0 (56.71)
Age = 4: 0 (34.0)
Age = 1: 0 (30.0)
Age = 5: 0 (22.0/2.0)
Occupation = 6: 1 (16.0/3.0) : 0 (3.0)
Number of Rules : 8

The results indicate that the attributes, which describe the classification, are the variables Age and Occupation. This means that variable Infected is more closely related to the variables Age and Occupation than the other variables and therefore in some Ages (Old people) and some Occupations (Farmers) there are more infected persons than in other categories. A possible explanation for these results is that the older people are more vulnerable and farmers have more probabilities to be bitten by mosquitoes. Table 4 presents the overall accuracy of the model computed from the training dataset and is equal to 97.13%.

Table 4. Stratified cross-validation.

Summary		
Correctly Classified Instances	305	97.1338%
Incorrectly Classified Instances	9	2.8662%
Kappa statistic		0.7278
Mean absolute error		0.0421
Root mean squared error		0.1526
Relative absolute error		42.229%
Root relative squared error		69.3346%
Coverage of cases (0.95 level)		99.0446%
Mean rel. region size (0.95 level)		55.8917%
Total Number of Instances		314

Table 5 presents that the worst performance based on the F-measure that combines precision and recall is for the class Infected and equals 74.3%, whereas the best performance is for the class Not Infected and equals 98.5%.

Table 5. Detailed Accuracy By Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,980	0,188	0,990	0,980	0,985	0,731	0,939	0,994	0
	0,813	0,020	0,684	0,813	0,743	0,731	0,939	0,607	1
Weighted Avg	0,971	0,179	0,974	0,971	0,972	0,731	0,939	0,974	

Table 6 presents the confusion matrix for the two classes, a:infected and b: Not infected.

Table 6. Confusion Matrix

a	b	<-- classified as
292	6	a=0
3	13	b=1

3.3 Clustering

The clustering step was performed using the k-means algorithm (SimpleKmeans in the context of WEKA). The number of clusters is set to 2, since the variable "Infected" was used to compute the accuracy of the clustering and inspect the impact of the Infected to the other variables.

Table 7. Clustering results. Variable "Infected" is used for assessing the clustering.

Final cluster centroids			
		Cluster#	
Attribute	Full Data	0	1
	(314)	(182)	(132)
Age	2	2	6
No_Children	2	2	2
Family_Members	2	4	2
Occupation	7	5	7
Education	1	2	1
Exp_Selfprot_Program	1	1	1

Table 7 shows the results of the clustering. The incorrectly clustered instances (farmers) are 36.94% based on variable "Infected". It is also evident from the cluster centroids that not infected, represented as cluster 0 in the results, contains households that have not persons infected by WNV compared to cluster 1 (persons infected by WNV).

3.4 Association rule mining

The Apriori algorithm (Agrawal et al., 1993) was used for finding association rules for our dataset. The algorithm was executed using a minimum support of 0.1 and a minimum confidence of 0.9, as parameters. WEKA produced a list of 14 rules (Table 8) with the support of the antecedent and the consequent (total number of items) at 0.1 minimum, and the confidence of the rule at 0.9 minimum (percentage of items in a 0 to 1 scale).

Table 8. The Apriori algorithm results based on the confidence metric.

Best rules found			
1.	Education=2 105	=>	Infected=0 105 <conf:(1)> lift:(1.05) lev:(0.02) [5] conv:(5.35)
2.	Education=2 Exp_Selfprot_Program=1 100	=>	Infected=0 100 <conf:(1)> lift:(1.05) lev:(0.02) [5] conv:(5.1)
3.	Occupation=7 90	=>	Infected=0 89 <conf:(0.99)> lift:(1.04) lev:(0.01) [3] conv:(2.29)
4.	Occupation=7 Exp_Selfprot_Program=1 87	=>	Infected=0 86 <conf:(0.99)> lift:(1.04) lev:(0.01) [3] conv:(2.22)
5.	Infected=0 Family_Members=2 81	=>	Exp_Selfprot_Program=1 79 <conf:(0.98)> lift:(1.05) lev:(0.01) [3] conv:(1.89)
6.	Infected=0 Education=1 98	=>	Exp_Selfprot_Program=1 95 <conf:(0.97)> lift:(1.04) lev:(0.01) [3] conv:(1.72)
7.	Occupation=7 90	=>	Exp_Selfprot_Program=1 87 <conf:(0.97)> lift:(1.04) lev:(0.01) [3] conv:(1.58)
8.	Infected=0 Occupation=7 89	=>	Exp_Selfprot_Program=1 86 <conf:(0.97)> lift:(1.04) lev:(0.01) [3] conv:(1.56)
9.	Exp_Selfprot_Program=1 292	=>	Infected=0 280 <conf:(0.96)> lift:(1.01) lev:(0.01) [2] conv:(1.14)
10.	Family_Members=2 90	=>	Exp_Selfprot_Program=1 86 <conf:(0.96)> lift:(1.03) lev:(0.01) [2] conv:(1.26)
11.	Occupation=7 90	=>	Infected=0 Exp_Selfprot_Program=1 86 <conf:(0.96)> lift:(1.07) lev:(0.02) [5] conv:(1.95)
12.	Education=2 105	=>	Exp_Selfprot_Program=1 100 <conf:(0.95)> lift:(1.02) lev:(0.01) [2] conv:(1.23)
13.	Infected=0 Education=2 105	=>	Exp_Selfprot_Program=1 100 <conf:(0.95)> lift:(1.02) lev:(0.01) [2] conv:(1.23)
14.	Education=2 105	=>	Infected=0 Exp_Selfprot_Program=1 100 <conf:(0.95)> lift:(1.07) lev:(0.02) [6] conv:(1.89)

The application of the Apriori algorithm for association provided some interesting outcomes for the attributes regarding the protection of households against the mosquitoes. Table 8 shows the association rules that can be discovered. There are of course some uninteresting rules, like rules 2 and 4. They present relatively known information since it is an expected or conforming relationship between variables. There are also a couple of symmetrical rules, since the antecedent element and the consequent element are interchanged. There is also a similar triad of rules, rules with the same element in antecedent and consequent but interchanged, such as rules 1, 3 and 7.

Summarizing the results from the classification, the clustering and the association rule mining methods we can conclude that:

(i) The attributes which best describe the classification are the variables Age and occupation. The attribute “Infected” is used as a class.

(ii) Using “Infected” as class attribute in clustering, namely if there is a person who was infected or nor, the results show that the two clusters have only two attributes with the same value (No_Children and Exp_Selfprot_Program).

(iii) In association rule mining, although there are some trivial rules, namely expected and previously known, like rules 2 and 4 that show that are mutually dependent, there are also rules like 1, 3, 8, 9 and 10, which offer a lot of actionability.

Overall, the infection by mosquitoes depends on Age, Occupation and Education and surprisingly not on the expenses for the protection against them.

4 Conclusions

The application of three different data mining techniques towards self-protection methods against mosquitoes is presented in this paper. The results show some interesting outcomes.

This study investigated the main demographic characteristics that affect the buying behavior of people towards the self – protection measures against mosquitoes. It showed that the amount consumers' spend for buying goods for self-protection against mosquitoes is mainly affected by the existence of children, age, number of family members, education and occupation. Hence this study supports the arguments of other study that socioeconomic conditions are associated with self-protection measures against mosquitoes. Furthermore, this study indicated that the people expenses towards self-protection methods against mosquitoes are significantly related to their infection from WNV. Besides, people who have been infected from WNV have different demographic profile in comparison with people who have not been infected.

According to the results of the current study, most of the people who spend less than 20 euro per month per household for self-protection against mosquitoes are mainly old (more than 61 years old, low educated, retiree, with two people in their family) and these people mainly consist of the population who are mainly infected by WNV.

Further investigation is still required since the results are based on only region. In the future we plan to extend the study to other regions.

References

1. Agrawal R., Imielinski, T. and Swami, A.N. (1993) Mining Association Rules between Sets of Items in Large Databases. In *Proc. of SIGMOD*, 207-216.

2. Agrawal, R. and Srikant, R. (1994) Fast algorithms for mining association rules. Proceedings of 20th International Conference on Very Large Data Bases (pp. 487-499).
3. Al-Dubai SA, Ganasegeran K, Mohanad Rahman A, Alshagga MA and Saif-Ali R. (2013) Factors affecting dengue fever knowledge, attitudes and practices among selected urban, semi-urban and rural communities in Malaysia. Southeast Asian J Trop Med Public Health.44(1):37–49. pmid:23682436.
4. Ansari MA, Sharma VP, Razdan RK and Mittal PK (1990) "Evaluation of certain mosquito repellents marketed in India" *Indian J Malariol* Vol 27 pp 57-64.
5. Barber, L., Schleier, J., and Peterson R. (2010) "Economic Cost Analysis of West Nile Virus Outbreak , Sacramento County, California, USA, 2005", Energing Infectious Diseases, Vol. 16, (3), March 2010, 480-485
6. Carney RM, Husted S., Jean C., Glaser C. and Kramer V., (2008) "Efficacy of aerial spraying of mosquito adulticide in reducing incidence of West Nile Virus, California 2005" Energing Infectious Diseases, Vol. 14, 747-754
7. Centers of Disease Control and Prevention (2015) *West Nile virus: clinical description* [http://www.cdc.gov/ncidod/dvbid/westnile/clinicians/clindesc.htm#diag (last access 12 September 2015]
8. Cunningham, S. J., and Holmes, G. (1999) Developing innovative applications in agriculture using data mining. In the Proceedings of the Southeast Asia Regional Computer Confederation Conference.
9. Curtis CF, Klines JD, Ijumba J, Callaghan A, Hill N and Karimzad MA (1989) "The relative efficacy of repellents against mosquito vectors of disease" *Med Vet Entomol* Vol 1 pp. 109-119.
10. Thitiprayoonwongse, D. Suriyapkol P. and Soonthornphisaj, N. (2012) "Data mining of dengue infection using decision tree," in Latest Advances in Information Science and Applications: The 12th WSEAS International Conference on Applied Computer Science, pp. 154–159.
11. García, E., Romero, C., Ventura, S., and Calders, T. (2007) Drawbacks and solutions of applying association rule mining in learning management systems. In Proceedings of the International Workshop on Applying Data Mining in e-Learning (ADML 2007), Crete, Greece (pp. 13-22).
12. Garcia-Rejon JE, Farfan-Ale JA, Ulloa A, Flores-Flores LF, Rosado-Paredes E, Baak-Baak C, Lorono-Pino MA, Fernandez-Salas I and Beaty BJ, (2008) Gonotrophic cycle estimate for Culex quinquefasciatus in Merida, Yucatan, Mexico. J Am Mosq Control Assoc. 24: 344-348.
13. He C., Hu X., Wang G., Zhao W., Sun D., Li Y., Chen C., Du J. and Wang S. (2014) "Eliminating *Plasmodium falciparum* in Hainan, China: a study on the use of baharioural change communication intervention to promote malaria prevention in mountain working populations", *Malaria Journal*, Vol. 13:273.
14. Hernandez-Romano J., Rodriguez M., Pando V., Torres-Monzon J., Alvarado-Delgado A. and Lecona Valera A. (2011) Conserved peptide sequences bind to actin and enolase on the surface of Plasmodium berghei ookinetes. Parasitology 138, 1341–1353 10.1017/S0031182011001296.

15. Itrat A, Khan A, Javaid S, Kamal M, Khan H, Javed S, Kalia S, Khan AH, Sethi MI and Jehan I. (2008) Knowledge, awareness and practices regarding dengue fever among the adult population of dengue Hit cosmopolitan. PLoS ONE. 3: 1-6.
16. Jones C.H., Benitez-Valladares D., Guillermo-May G., Dzul-Manzanilla F., Che-Mendoza A., Barrera-Perez M., Selem-Salas C., Chable-Santos J., Sommerfeld J., Kroeger A., O'Dempsey T., Medina-Barreiro A. and Manrique-Saide R. (2014) "Use and acceptance of long lasting insecticidal net screens for dengue prevention in Acapulco, Guerrero, Mexico", *BMC Public Health,* Vol. 14:846
17. Kantardzic, M. (2003) Data Mining: Concepts, Models, Methods, and Algorithms. New York, NY: John Wiley & Sons.
18. Kaufmann, L. and Rousseeuw, P.J. (1990) Finding Groups in Data: An Introduction to Cluster Analysis, New York, John Wiley & Sons.
19. Koenraadt C., Tuiten W., Sithiprasasna R., Kijchalao U., Jones J. and Scott T. (2006) "Dengue Knowledge and Practices and their Impact on Aedes Aegypti Populations in Kamhaeng Phet, Thailand", *Am. J. Trop Med. Hyg.* vol. 74, No, 4, pp. 692-700.
20. Liu, B. and Hsu, W. (1996) Post-Analysis of Learned Rules. Proceedings of National Conference on Artificial Intelligence. Portland, Oregon, USA, (pp. 828–834).
21. Liu, B., Hsu, W., Chen, S. and Ma, Y. (2000) Analyzing the Subjective Interestingness of Association Rules. IEEE Intelligent Systems, 15(5), 47–55.
22. Lorono-Pino M. A., Chan-Dzu Y.N., Zapata-Gil R., Carrillo-Solis C., Uitz-Mena A., Garcia-Rejon J.E., Keffe T.J., Beaty B.J. and Eisen L. (2014) "Household use of insecticide consumer products in a dengue-endemic area in Mexico" *Tropical Medicine and International Health,* Vol. 19, No. 10, pp.1267-1275.
23. MacQueen, J. (1967) Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth berkeley symposium on mathematical statistics and probability, (pp. 281–297). California, USA.
24. Mayxay M, Chotivanich K, Pukrittayakamee S, Newton P, Looareesuwan S and White NJ. (2001) Contribution of humoral immunity to the therapeutic response in falciparum malaria. Am J Trop Med Hyg. 65(6):918–23. Epub 2002/01/17. pmid:11791999
25. Minaei-Bidgoli, B., Tan, P-N. & Punch, W.F. (2004) Mining Interesting Contrast Rules for a Web-based Educational System. Proceedings of Int. Conf. on Machine Learning Applications, Louisville, USA 2004 (pp. 320- 327).
26. Mostashari F, Kulldorff M, Hartman JJ, Miller JR and Kulasekera V (2003) Dead bird clustering: A potential early warning system for West Nile virus activity. Emerg Infect Dis 9: 641–646.F. MostashariM. KulldorffJJ HartmanJR MillerV. KulasekeraDead bird clustering: A potential early warning system for West Nile virus activity.Emerg Infect Dis20039641646
27. Naing C, Ren WY, Man CY, Fern KP, Qiqi C, Ning CN and Ee CW, (2011), Awareness of dengue and practice of dengue control among the semi-urban community: a cross-sectional survey. J Community Health. 36: 1044-1049. 10.1007/s10900-011-9407-1.

28. Nalwanga E and Ssempebwa JC, (2011) Knowledge and practices of in-home pesticide use: a community survey in Uganda. J Environ Publ Health. 2011, 230894-Accessed at URL: http://www.hindawi.com/journals/jeph/2011/230894 on 20th November 2011
29. Pandit N, Patel Y and Bhavsar B. (2010) "Awareness and practice about preventive method against mosquito bite in Gujarat". *Healthline* Vol. 1: 16–20.
30. Raghavendra K, Baril TK, Reddy BPN, Sharma P, Dash AP (2011) "Malaria vector control: from past to future" *Parasitol Res.* Vol. 108 pp. 757-779.
31. Rozendaal JA (1997) *Vector Control: methods for use by individuals and communities*, World Health Organization, Geneva.
32. Shuaib F, Todd D, Campbell-Stennett D, Ehiri J, Jolly PE (2010) Knowledge, attitudes and practices regarding dengue infection in Westmoreland, Jamaica. West Indian Med J 59: 139–146.
33. Syed M, Saleem T, Syeda UR, Habib M, Zahid R, Bashir A, Rabbani M, Khalid M, Iqbal A, Rao EZ, Shujja Ur R and Saleem S. (2010) Knowledge, attitudes and practices regarding dengue fever among adults of high and low socioeconomic groups. J Pak Med Assoc. 60: 243-247.
34. Timmis, J., Andrews, P., and Hart, E. (2010) On artificial immune systems and swarm intelligence. Swarm Intelligence, 4(4):247–273.
35. Tyagi P, Roy A. and Malhotra M.S. (2005) "Knowledge, awareness and practices towards malaria in communities of rural, semi-rural and bordering areas of east Delhi (India)" *J. Veet Bonne Dis.* Vol. 42, pp.30-35, March 2005
36. Van Benthem BH, Khantikul N, Panart K, Kessels PJ and Somboon P (2002) Knowledge and use of prevention measures related to dengue in northern Thailand. Trop Med Int Health 7: 993–1000.
37. Witten, I. and Frank, E., (2005) Data Mining Practical Machine Learning Tools and Techniques, San Francisco: Morgan Kaufmann.
38. World Health Organization, (1998) The world health report 1998—life in the 21st century: a vision for all. http://www.who.int/whr/1998/en/ (accessed Oct 2016).