A Rule-Based Quality Analytics System for the Global Wine Industry

Carmen K. H. Lee, School of Business, Singapore University of Social Sciences, Singapore

Kris M. Y. Law, School of Engineering, Deakin University, Australia

Andrew W. H. Ip, University of Saskatchewan, Canada

https://orcid.org/0000-0001-6609-0713

ABSTRACT

The global wine-making industry has faced challenges due to the increasing demands of consumers, particularly in emerging markets such as China, Brazil, India, and Russia. Controlling the quality during wine production is one of the key challenges faced by global winemakers to produce wine with appropriate sensorial properties tailored to specific markets. The wine production quality is constituted from a number of environmental factors such as climate, soil, and temperature, which affect the sensorial properties and the overall quality. This paper proposed a rule-based quality analytics system (RBQAS) to capture physicochemical data during wine production and to investigate the hidden patterns from the data for quality prediction. It consists of IoT for data capture on a real-time basis, followed by association rule mining to identify relationships between sensorial and physicochemical properties of wine.

KEYWORDS

Association Rules, Data Mining, Global Wine Industry, Physiochemical Property, Predictive Analytics, Quality Management, Sensorial Property, Wine Production

1. INTRODUCTION

With customers nowadays having higher expectations on products, firms are urged to improve their product quality with continuous effort, in order to remain competitive in the market. The foremost is to ensure the conformance of products with given requirements and standards and this conformance can be achieved through systematic measurement and control to detect special causes of variation. Jabnoun (2002) proposed five processes involved in quality assurance: setting the standards, providing the inputs that enable workers to conform to standards, measuring performance, analyzing the measured performance data, and taking corrective actions. However, these quality assurance processes tend to be reactive. Problems are corrected only when the defects are identified. A more proactive approach is needed to predict the quality and take corrective actions earlier before the products with poor quality are produced.

The global wine industry is forced to be more aware of the quality assessment and testing, due to the increasing demand. Wine making is sophisticated, which is determined by a number of factors such as climate, soil and temperature. These environmental factors lead to the physicochemical properties of wine, such as pH, density and acidity, affecting the sensorial properties and the overall quality of wine.

Without knowledge support driven from historical data, decision makers are not effective in determining optimal winemaking conditions which have a great impact on both the physicochemical and sensorial properties of wine.

To optimize the cost and quality, it is essential to have effective determination on the winemaking conditions, as the input resources are highly related to the production cost. Furthermore, the lack of technological resources makes it difficult for the industry to classify the wines based on the chemical analysis as it takes lot of time which imposes impact on the financial.

Faced with the challenges from the global market, in this paper, a rule-based quality analytics system (RBQAS) for the global wine industry is proposed to capture physicochemical data during wine production and to investigate the hidden patterns from the data for quality prediction. This system is developed with the motivation and mission to provide wine manufacturers a better structured and managed quality analytics system, which can provide global wine makers a more standardized quality.

The proposed RBQAS consists of data capture on a real-time basis, using the latest IoT technology, followed by association rule mining to identify relationships between sensorial and physicochemical properties of wine. Association rules are mined on a quality data set for quality analytics of wine. It relates combinations of quality attributes (wine quality) and subsets of independent attributes (the physiochemical properties of wine). The rules represent valuable knowledge to predict and improve the wine quality.

Nevertheless, there are two challenges faced by the wine manufacturers when applying association rule mining for quality management. The first challenge is the extremely large number of rules obtained, it becomes impractical for wine manufacturers to identify critical rules for quality prediction. The second problem is that the discovered rules without validation could be with poor generalization potential. These highlight the importance of rule reduction and rule validation in quality analytics.

The previous works using data mining approaches with different algorithms mainly focus on the prediction of wine quality, however, do not show the causal effects and correlations of the parameters.

Accordingly, this paper addresses the following three research questions:

1. How to conduct a rule-based quality analytics to discover the relationship between sensiorial and physicochecmial properties of wine?
2. How to reduce the amount of association rules for the ease of decision making in quality analytics?
3. How to identify and remove rules with poor generalization potential so that the rules could be general enough for applications on a global scale?

The rule-based algorithm proposed in this paper advances the existing research, by offering a better understanding on the causal effects of the parameters on wine quality. Specifically, the proposed algorithm supports both descriptive and predictive analytics, wine manufacturers are provided with decision support in determining the appropriate chemical parameters and corrective measures for better wine quality. Furthermore, the knowledge discovered by the RBQAS can serve as an additional supplement to wine manufacturers' experience and is feasible to be generalized for wider applications in the wine industry.

2. LITERATURE REVIEW

2.1. Wine Production

The global wine making industry has faced challenges due to the increasing demands of consumers, particularly in emerging markets such as China, Brazil, India and Russia (Gilinsky et al., 2016). Controlling the quality during wine production is critical to the acceptance and preferences of wine (Niimi et al., 2017). The key challenge faced by winemakers is to produce wine with appropriate sensorial properties tailored to specific market segments as liking of wine based on sensory perception

is not homogenous within a population (Biasoto et al., 2014). Nevertheless, the relationships between the intricate sensory and chemical components contained in wine are inherently complex (Bastian, et al., 2010). The environmental issues affect not only the quality of wine production but also the the impact on ecosystems, while nowadays sustainability has been called for in the global market. The main focus of sustainable business is to preserve the environment to guarantee the wellbeing of future generations (Ouvrard, Jasimuddin & Spiga, 2020).

Traditionally, the determination of production parameters heavily relies on the experience and knowledge of winemakers, where trial and error processes are usually adopted to customize for different customers. This unveils a lack of scientific approaches to understanding the relationships between production parameters and the resultant quality of the wine. The lack of technological resources makes it difficult for the industry to classify the wines based on the chemical analysis as it takes lot of time which imposes impact on the financial.

2.2. Data Mining for Quality Management

Product quality can be improved through the discovery of contributing parameters from the hidden knowledge (Ur-Rahman and Harding, 2012). Data mining techniques are the promising tools, which help firms leverage their data more effectively and extract hidden knowledge for quality prediction (Köksal et al., 2011). In particular, association rule mining is useful in identifying the hidden relationships between items that occur together (Altuntas & Selim, 2012). The Apriori Algorithm proposed by Agrawal and Srikant (1994) is a classical algorithm for association rule mining, which employs a downward closure property (Coenen, Leng & Ahmed, 2004; Song & Rajasekaran, 2006). A rule is considered useful if it meets user-defined threshold values. Both threshold values for support and confidence which are defined by users will thus affect the quality of association rules (Demiriz et al., 2011; Lim et al., 2012). Thus, it is important to ensure that user-specified minimum support and confidence are appropriate to generate significant rules.

2.3. Association Rule Mining in Big Data

Advanced technologies, such as IoT and mobile technologies, generate an enormous volume of data in various formats rapidly including Radio Frequency Identification and Wireless Sensor Networks (Xu et al., 2014; Tsang et al., 2017; Arpan et al, 2017; Ng et al., 2018). One of the most important outcomes of these emerging technologies is the creation of an unprecedented amount of data (Gubbi et al., 2013). Big data, characterized with high volume, variety and velocity, increasingly drives decision making and is transforming the landscape of business intelligence. Big data analytics that discover insights from evidences have an increasing demand for computing efficiency, knowledge discovery, problem solving, and prediction.

Association rule mining therefore provides a useful mechanism for discovering the correlations among quality related attributes for prediction. The applications of association rule mining include the root cause analysis of anomalies to improve service and repair in the automotive domain (Chougule et al., 2011), the discovery of defect patterns for defect prediction (Chang et al., 2009; Lee et al., 2013), optimization (Khatwani and Srivastava, 2017), and the discovery of the relationship between predefined events in multiple time series, so as to explain failures in the hot-dip galvanizing of steel lines (Martínez-de-Pisón et al., 2012).

In the era of big data, difficulties arise when association rule mining is applied for quality management. Most algorithms accomplish the association tasks by finding frequent itemsets and generating useful rules based on the frequent itemsets (Hand, Mannila, & Smyth, 2001). The first difficulty is that the number of rules obtained could be extremely large, so it becomes impractical for decision makers to identify critical rules for quality prediction. The second problem is that the discovered rules without validation could be with poor generalization potential. These highlight the importance of rule reduction and rule validation in predictive quality analytics.

There is previous work on using various constraints to reduce the number of rules. Constraints on support were used for pruning the number of candidate itemsets generated (Wang, He & Han, 2003). Item constraints were incorporated to include or exclude certain items in the rules in Srikant, Vu and Agrawal (1997). Constraints imposed on the antecedent and consequents of the rules to be mined were used in Ng et al. (1998). Nevertheless, previous work focused on mining association rules on the entire data set, without validation on an independent data set except in Ordonez (2006). To the best of our knowledge, reducing the number of association rules with constraints and validation on an independent set has not previously been studied in the context of predictive quality analytics.

3. THE RULE-BASED QUALITY ANALYTICS SYSTEM

Figure 1 shows the architecture of the proposed RBQAS. There are three modules, namely Rule-based Knowledge Discovery Module (RKDM), Rule Validation Module (RVM) and IoT Production Monitoring Module (IPMM).

The RKDM embeds an association rule mining algorithm to analyse the data related to wine quality. To facilitate the knowledge discovery for wine quality management, the data is classified into two types: sensorial data and physicochemical data. The objective of the association rule mining algorithm is to discover the hidden relationship between the physicochemical and sensorial properties of the wine. In addition, constraints are introduced in the algorithm to reduce the number of rules obtained. This algorithm is anticipated to advance the previous works on incorporating constraints for association rules, including Wang, He & Han (2003), and Ng et al. (1998).

The RVM is responsible for the cross validation process of the association rules obtained from the RKDM. The validation process is important before the rules can be deployed in an actual wine production environment. Without validation, rules may not be general enough for deployment. Train and test approaches are designed for the cross validation. The dataset is partitioned into two subsets: training set and testing set. Only the training set is used in the RKDM for generating the association rules. The training rules obtained are then evaluated using the testing set. The process is repeated multiple times and the output is a set of association rules that are considered valid.

The IPMM is about the deployment of the valid association rules in an actual wine production environment. The rules suggest the optimal physicochemical conditions for wine production and their relationships with the resultant wine quality. Wine manufacturers currently monitor and control the wine production conditions based on their experience so as to ensure that the resultant quality of wine is desirable. The knowledge discovered by the RBQAS can serve as an additional supplement for wine manufacturers' experience. To make good use of the rules, it is critical to monitor the wine production condition such as the physiochemical parameters, so that appropriate adjustment of the parameters can be done when necessary. IoT technologies are proposed for real-time capture the data for monitoring and quality control purposes. Sensors such as pH sensors, water quality sensors and temperature sensors in the IPMM are suggested for capturing the physicochemical data and monitoring the production. Data captured by the sensors, together with the sensory evaluation by wine experts, becomes part of the new input for the RKDM.

Details of each module are presented in the following sections.

3.1. Rule-Based Knowledge Discovery Module

The core part of the RKDM is the constrained-rule mining algorithm, based on Apriori algorithm, which discovers the relationship between the physicochemical and sensorial properties of the wine. The input data has to be divided into two types: sensorial data and physicochemical data. Examples of sensorial data are the texture, sweetness, and overall rating of the wine based on the judgment of wine experts. Examples of physicochemical data include the acidity, pH value, density and alcohol level of the wine. Considered that the sensorial data and physicochemical data are numeric data but association rule mining can only handle categorical data, data transformation is performed to

Figure 1. The proposed RBQAS
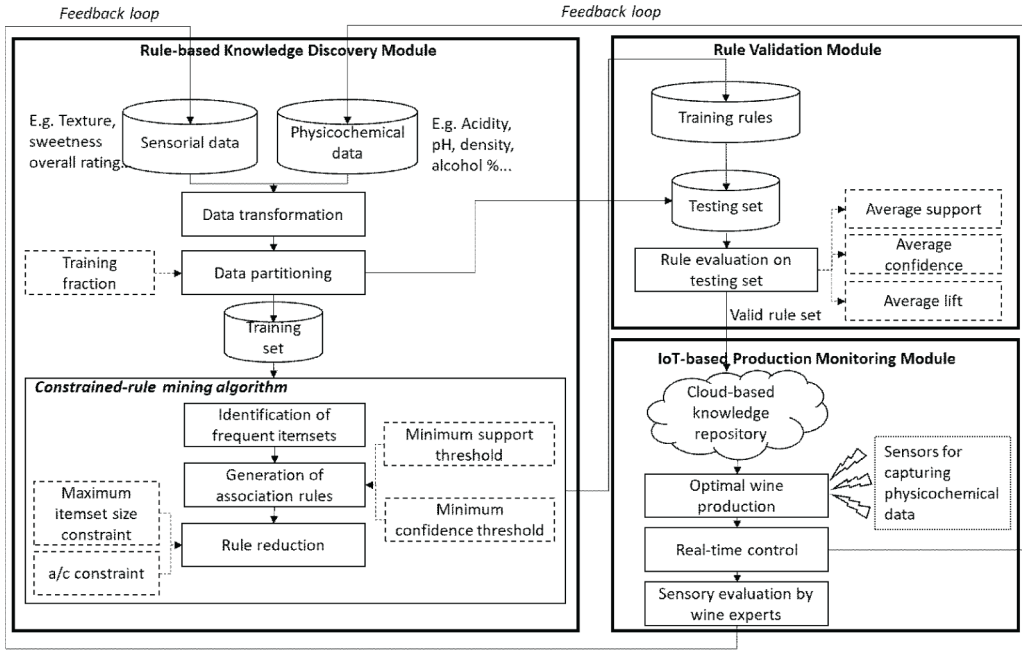
convert the numeric data into cateogircal data. For instance, the pH value ranged from 1 to 14 could be transformed into such cateogries as "low pH value", "medium pH value" and "high pH value".

In association rule mining, each category associated with the attributes are viewed as an item. Let $I = \{ i_1, i_2, ..., i_n \}$ be a set of items. An itemset X is a subset of I. X containing k items is called a k-itemset. For example, {low pH value} is an 1-itemset while {low pH value, high quality} is a 2-itemset.

Let $D = \{ T_1, T_2, ..., T_m \}$ be a transaction database, where T_i ($i \hat{I}$ [1...m]) is a transaction containing a set of items in I. An association rule is generated using frequent itemsets which are defined as itemsets that frequently appear in D. Thus, for each itemset X, its probability of appearance in D has to be calculated and is termed as the support of X.

$s(X)$, the support of X, is the number of transactions containing X in D. It is the probability of appearance of X in D, i.e. $P(X) = s(X)$. If $s(X)$ is larger than or equal to a user-specified minimum support threshold δ, then X is a frequent itemset and can be used to generate association rules which consist of antecedent and consequent parts. For example, if {low pH value, high quality} is a frequent itemset, there are two association rules generated:

- If the wine has low pH value, then the wine has high quality;
- If the wine has high quality, then the wine has low pH value.

The confidence of the rules represent the strength of the association of the items listed in the rules. Rule confidence $c(X \Rightarrow Y)$ is the conditional probability of appearance of Y given X. Consequently, $c(X \Rightarrow Y) = P(Y|X) = P(X \cup Y)/P(X) = s(X \cup Y)/s(X)$. The rule is considered reliable if its confidence is greater than or equal to a user-specified minimum confidence threshold α.

Another measure called lift, $l(X \Rightarrow Y)$, is also used to quantify the relationship between X and Y. It is defined as $l(X \Rightarrow Y) = c(X \Rightarrow Y)/s(Y)$. In general, a lift value greater than 1.0 implies that X and Y depend on each other.

To speed up the computation time, the Apriori algorithm generates association rules in two phases: finding frequent itemsets using δ (Phase 1) and generating rules using α and the lift threshold τ (Phase 2). In addition, to eliminate irrelevant rules, two search constraints are designed and incorporated in the aglorithms. The first constraint is called the maximum itemset size constraint. It is used to limit the size of the itemsets to be used in rule generation. The second constraint is called the antecedent-consequent (a/c) constraint. It is used to define whether an item can appear in the "antecedent" or the "consequent" parts of an association rules. Details of each constraint are given below.

3.1.1. Maximum Itemset Size Constraint

The association rules are generated using the frequent itemsets. Thus, the number of rules increases with the number of frequent itemsets. Given n unique items in I, the number of possible association rules can be up to $3n$-$2n+1+1$. This constraint restricts the size of frequent itemsets up to a maximum size g, terminating the search process of frequent itemsets with size larger than g. It is applied in Phase 1 of the algorithm where the itemsets are generated. It is simple but effective in reducing the number of rules. Another reason supporting the incorporation of this constraint is that the associations among too many items may be irrelevant for quality prediction when the itemset is large. Though this constraint can also be applied in Phase 2, it is suggested applying it as early as possible (i.e. in Phase 1) to improve the computation efficiency.

3.1.2. A/C Constraint

The a/c constraint is used to reduce the number of rules by specifying in which part of the rules the items should appear. There are three possible conditions for an item to appear in a rule. First, an item can only appear in the antecedent of a rule. Second, an item can only appear in the consequent of a rule. Third, an item can appear in either antecedent or consequent of a rule. For quality management, items related to quality should appear in the consequent of a rule while other items affecting the resultant quality should appear in the antecedent of a rule. This constraint eliminates all the rules containing items that do not appear in the correct place of the rules. It plays an important role in pruning rules that are not correspond to the data mining goal.

3.2. Rule Validation Module

In the RVM, train and test approaches are used to validate the rules. Data partitioning is performed to devide D into two independent subsets: training set D_{train} and testing set D_{test}. A traning fraction μ is set to define the size of D_{train}. In the RBQAS, only D_{train} is sent to the RKDM to generate association rules. The rules generated based on the training set are labelled as training rules R_{train}. After that, R_{train} is validated on $D_{test.}$ Using the D_{test}, each of the training rule $X \Rightarrow Y$ is evaluated in terms of $s(X \Rightarrow Y)$, $c(X \Rightarrow Y)$ and $l(X \Rightarrow Y)$. The rule will be discarded if its support, confidence and lift values are smaller than the thresholds δ, α, τ, respectively. Otherwise, it will be kept as a testing rule R_{test}.

This process is repeated t times on independent subsets. In each iteration, all the testing rules obtained are labeled as a set of rules R_t. Only rules that appear in all the R_t are considered valid. Let $X \Rightarrow Y$ be a valid rule appearing in all R_t. The strength of association between X and Y is evaluated in terms of the average support, confidence and lift values as (1)-(3):

$$s\left(X \Rightarrow Y\right) = t^{-1} \sum_{j=1}^{t} s\left(X \Rightarrow Y, R_j\right) \tag{1}$$

$$c\left(X \Rightarrow Y\right) = t^{-1} \sum_{j=1}^{t} c\left(X \Rightarrow Y, R_j\right) \tag{2}$$

$$l\left(X \Rightarrow Y\right) = t^{-1} \sum_{j=1}^{t} l\left(X \Rightarrow Y, R_j\right) \tag{3}$$

3.3. IoT-Based Production Monitoring Module

In the IPMM module, the IoT technology stack consists of the device layer, the connectivity layer and the IoT cloud layer. In the device layer, sensors are deployed to monitor the physiochemical parameters during production on a real-time basis. Certain software and programmes are flashed on sensor nodes to operate the requested functionalities, such as data measurement intervals. The data captured by the devices, such as pH sensors, water quality sensors and temperature sensors, is transmitted to the IoT cloud layer through wireless communication protocols in the connectivity layer. In the IoT cloud layer, a platform is established to register the sensor nodes, manage the devices, and provide both front-end and back-end development. The captured data is uploaded to a cloud database so that operators are able to access the data for real-time monitoring and control. Any changes in the production can be captured by the physical devices and loaded to the cloud database under the IoT environment. On the other hand, the resultant sensorial properties of the wine produced are evaluated by the wine experts. For instance, the overall quality of the wine can be rated by the experts using a scale from 0 to 10.

Both the sensorial data recorded by the wine experts, and the physicochemical data in the cloud database become new input to the RKDM for data mining, forming a closed-loop for feedback. With a larger data size for data mining, more knowledge can be recursively discovered. This provides the IoTQAS with a learning capability which can enhance the quality of decision rules over time.

4. VERIFICATION OF THE CONSTRIANED RULE LEARMNING ALGORITHM

The data set created by Cortez et al. (2009) is used to illustrate how the relationship between physicochemical and sensorial data could be mined using the constrained-rule learning algorithm. It contains a total of 4,898 records related to white variants of the Portuguese Vinho Verde wine. There are 12 attributes describing the physicochemical and sensorial properties of each observation. Table 1 shows the attributes involved in the data set and the range of each attribute. In particular, the sensorial data "Quality (Q)" is the median of at least three evaluations made by wine experts, each of which was asked to grade the wine quality in a scale that ranges from 0 (very bad) to 10 (very excellent) using blind tastes. According to the proposed RBQAS, we assumed that those attributes could be captured by sensors within the IoT network.

As shown in Table 1, the attributes in the data set are in numeric forms. In order to apply the Apriori algorithm, the numeric attributes have to be transformed into categorical forms. Specifically, they are discretized into intervals and each interval is considered as an item for association rule mining. For instance, FA is discretized into four different intervals: {FA<6.2}, {6.2<=FA<7}, {7<=FA<8.1} and {FA>=8.1}; VA is discretized into five intervals: {VA<0.19}, {0.19<=VA<0.26}, {0.26<=VA<0.35}, {0.35<=VA<0.51} and {VA>=0.51}. Similarly, Q is discretized into {Quality=Bad}, {Quality=Average}, {Quality=Satisfactory}, {Quality=Good}, {Quality=Very Good}, {Quality=Excellent} and {Quality=Very Excellent}". After data discretization, there are 57 items in the data set.

Table 1. Attributes in the data set

Type	Symbol	Attribute	Range
Physicochemical data	FA	Fixed acidity (g(tartaric acid)/dm^3)	3.8-14.2
	VA	Volatile acidity (g(acetic acid)/dm^3)	0.1-1.1
	CA	Citric acid (g/dm^3)	0-1.7
	RS	Residual sugar (g/dm^3)	0.6-65.8
	CL	Chlorides (g(sodium chloride)/dm^3)	0.01-0.35
	FSO2	Free sulfur dioxide (mg/dm^3)	2-289
	TSO2	Total sulfur dioxide (mg/dm^3)	9-440
	D	Density (g/cm^3)	0.987-1.039
	pH	pH value	2.7-3.8
	SO4	Sulphates (g(potassium sulphate)/dm^3)	0.2-1.1
	A	Alcohol (vol.%)	8.0-14.2
Sensorial data	Q	Quality	3-9

The verification of the constrained-rule learning algorithm is done in two parts. In the first part, the performance of the two constraints on rule reduction is assessed. To test the impact of the maximum itemet size constraint on the number of rules, the a/c constraint is excluded and the number of maximium itemset is set from 2 to 8. The number of association rules obtained is then compared in each scenario. Next, to test the impact of the a/c constraint on the number of rules, the experiments are repeated but with the a/c constraint included. The a/c constraint design depends on the data mining objective. The objective of the algorithm is to help the wine manufacturers identify the appropriate wine production conditions for controlling the quality of wine. Therefore, in the setting of the a/c constraint, only the 11 physicochemcial attributes will be allowed to appear in the antecedent part of an association rule whilst the sensorial attribute Q will only be allowed to appear in the consequent part of the rule. In addition, different parameter settings for δ, α, τ are tested and compared. δ is tested at 1%, 2% and 5% while α is set to be larger than 50%. τ is set at 1 as it generally implies that the items depend on each other.

In the second part of the verification, the quality of the rules generated by the algorithm is assessed using the train and test approaches. Firstly, different training fractions are tested. To ensure that there are sufficient records to generate association rules, the training fraction is set to be larger than 50%. It means that at least half of the records will be used for association rule mining. Under each training fraction setting, the training and test approaches are repeated 5 times. Only rules that are common in all the five testing sets are considered valid and their average values of support, confidence and lift are reported.

Details of each part of the verification are discussed in the following sections.

4.1. Assessing the Impact of Constraints on Rule Reduction

In this section, each constraint is studied individually to examine its relative importance with respect to the number of association rules under different parameter settings. Figure 2 shows the number of rules under different parameter settings. Obviously the number of rules decreases with α. This is because more rules would be deemed useful if the minimum rule confidence threshold is lower. On the other hand, it is found that the number of rules are the largest when the itemsets contain 5 itemsets. This shows that frequent 5-itemsets are dominant. There are not many frequent itemsets containing more than 5 itemsets and thus the number of rules declines when the size of itemsets is

Figure 2. Distribution of different sizes of rules
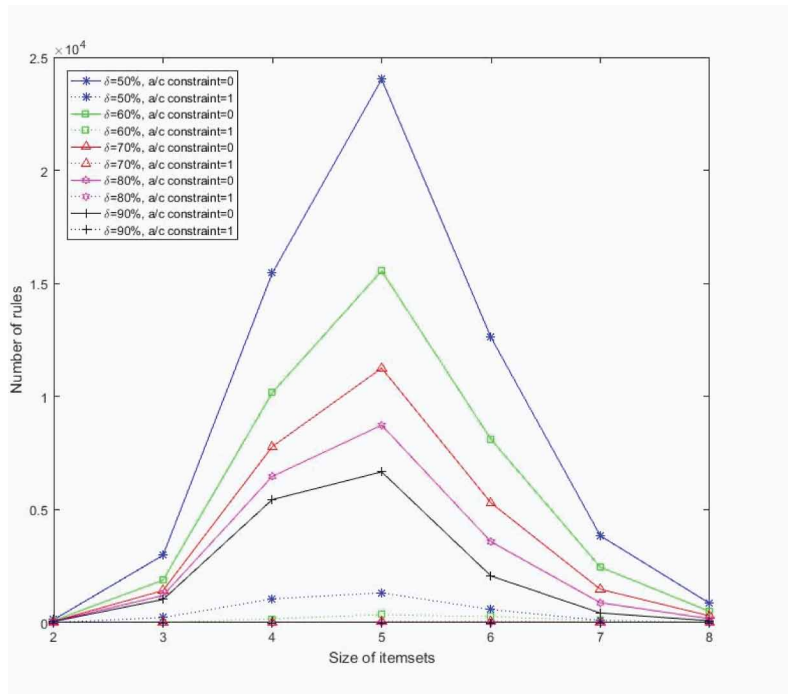
6 or above. With the introduction of '*a/c constraint*', the number of rules is greatly reduced as only rules having Q appearing in the consequent part and other items appearing in the antecedent part are kept. This implies that the impact of the a/c constraint on the reduction of rules becomes more significant when the size of itemsets increases.

In association rule mining, careful consideration for choosing appropriate δ and α is required. The number of rules increases when δ decreases. This is because, with a smaller δ, more itemsets can be considered frequent itemsets that are used to generate rules. On the other hand, the number of rules decreases when α increases. Only rules with confidence greater than or equal to α are considered useful. Therefore, when α decreases, more rules are obtained despite the fact that they might not be very reliable or meaningful. Tables 2-4 compare the number of rules under different values of δ and α. With the introduction of the two constraints, α should not be set greater than 80%, 70% and 60% when δ is set at 1%, 2% and 5%, respectively. Otherwise, no rules could be obtained. These findings illustrate the trade-off between δ and α.

In addition, when α is fixed, larger frequent itemsets are formed when δ decreases. For instance, from Table 2, frequent itemsets can be generated up to $g=8$ at $\delta = 1\%$. However, when δ is increased to 2%, frequent itemsets can be generated up to $g=7$ as shown in Table 3. Similarly, when δ is increased to 5%, frequent itemsets can be generated up to $g=5$ as shown in Table 4.

4.2. Evaluating Association Rules Using Train and Test Approaches

To validate the rules, different training fractions μ are set. Under each training fraction, the experiments are repeated five times. Only rules which are common in all the five testing sets are considered valid and their average values of support, confidence and lift are reported. Table 5 summarizes the numbers of rules in training and testing sets at $\mu = 50\%$, 60%, 70%, 80% and 90%, given $\delta = 1\%$, $\alpha = 50\%$, $\tau = 1$. It is observed that the change in the number of rules is not significant in the five experiments

Table 2. Impact of constraints on the number of rules at δ = 1%

	(a) δ =1%, α=50%, τ =1		(b) δ =1%, α=60%, τ =1		(c) δ =1%, α=70%, τ =1		(d) δ =1%, α=80%, τ =1		(e) δ =1%, α=90%, τ =1	
	Without a/c constraint	With a/c constraint	Without a/c constraint	With a/c constraint	Without a/c constraint	With a/c constraint	Without a/c constraint	With a/c constraint	Without a/c constraint	With a/c constraint
g=2	134	5	89	0	68	0	61	0	54	0
g=3	3114	230	1968	12	1490	1	1260	0	1079	0
g=4	18574	1282	12141	183	9276	14	7729	0	6523	0
g=5	42601	2603	27695	536	20531	54	16461	0	13190	0
g=6	55213	3184	35801	804	25827	102	20040	2	15254	0
g=7	59056	3287	38257	889	27304	126	20920	4	15681	0
g=8	59910	3298	38753	900	27602	130	21092	4	15767	0

Table 3. Impact of constraints on the number of rules at δ = 2%

	(a) δ =2%, α=50%, τ =1		(b) δ =2%, α=60%, τ =1		(c) δ =2%, α=70%, τ =1		(d) δ =2%, α=80%, τ =1	
	Without a/c constraint	With a/c constraint	Without a/c constraint	With a/c constraint	Without a/c constraint	With a/c constraint	Without a/c constraint	With a/c constraint
g=2	131	5	88	0	67	0	61	0
g=3	2165	133	1441	4	1120	1	967	0
g=4	8294	440	5652	40	4399	3	3738	0
g=5	13137	664	8863	96	6643	5	5419	0
g=6	14279	707	9623	123	7121	6	5714	0
g=7	14369	710	9685	126	7259	6	5734	0

Table 4. Impact of constraints on the number of rules at δ = 5%

	(a) δ =5%, α=50%, τ =1		(b) δ =5%, α=60%, τ =1		(c) δ =5%, α=70%, τ =1	
	Without a/c constraint	With a/c constraint	Without a/c constraint	With a/c constraint	Without a/c constraint	With a/c constraint
g=2	112	5	74	0	59	0
g=3	867	31	654	1	541	0
g=4	1710	57	1265	3	991	0
g=5	1881	62	1376	4	1052	0

and the average number of rules is calculated. Our results show that setting μ at 60% can obtain the largest number of testing rules.

On the other hand, the five different testing sets contain rules that might be in common or different. Only the common rules are considered valid in this study. Table 6 summarizes the common testing rules. When $\mu = 50\%$, 60%, 70%, 80% and 90%, the numbers of valid rules are 982, 1018, 840, 684, and 618, respectively. Only if $\mu = 80\%$ or 90%, there are rules with average confidence greater than or equal to 80%. Table 7 shows some selected rules with average confidence above 75%:

Table 5. Number of rules in training and testing sets under different training fractions

	Experiment 1		Experiment 2		Experiment 3		Experiment 4		Experiment 5		Average	
	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
$\mu = 50\%$	3596	1798	3806	1964	3896	1834	3336	1934	3940	1992	3714.8	1904.4
$\mu = 60\%$	3448	1896	3600	1924	3550	1890	3662	2028	3500	1812	3552	1910
$\mu = 70\%$	3488	1704	2930	1980	3412	1850	3366	1978	3096	1976	3258.4	1897.6
$\mu = 80\%$	3328	1718	3246	1970	3252	1714	3264	1968	3106	1898	3239.2	1853.6
$\mu = 90\%$	3142	1876	3118	1926	3200	1738	3206	1930	3214	1930	3176	1880

- The first rule implies that the quality of the wine woud be average if the chlorides level is at least $0.056g$(sodium chloride)/dm^3 but under $0.117g$(sodium chloride)/dm^3, volatile acidity is at least $0.26g$(acetic acid)/dm^3) but under $0.35g$(acetic acid)/dm^3, the pH value is at least 3.13 but below 3.22 and free sulfur dioxide amount is at least 40mg/dm^3 but less than 54mg/dm^3;

- The second rule implies that the quality of the wine woud be average if the chlorides level is at least $0.056g$(sodium chloride)/dm^3 but under $0.117g$(sodium chloride)/dm^3, volatile acidity is at least $0.35g$(acetic acid)/dm^3) but under $0.51g$(acetic acid)/dm^3, and citric acid is at least $0.46g$/dm^3;

- The third rule implies that the quality of the wine would be satisfactory if the sulphates level is at least $0.44g$(potassium sulphate)/dm^3 but under $0.59g$(potassium sulphate)/dm^3, free sulfur dioxide amount is at least 20mg/dm^3 but less than 29mg/dm3, the chlorides level is under $0.04g$(sodium chloride)/dm^3, and residual sugar under $3.15g$/dm^3.

5. DISCUSSION

5.1. Comparing Solutions With Existing Work

Many previous studies on wine quality focused on the accuracies of the classification of wine quality based on chemical parameters (e.g. Ribeiro et al., 2009; Appalasamy et al., 2012; Aich et al., 2018). Ribeiro et al. (2009) used decision trees, Artificial Neural Networks and linear regression to predict the organoleptic parameters, such as colour, foam, flabour, and savour, from the chemical parameters in the wine vinification process. Appalasamy et al. (2012) applied two classification algorithms, Decision tree and Naïve Bayes, to predict the quality of wine. The performance of the algorithms was compared in terms of accuracy rate. In a similar vein, Aich et al. (2018) compared the accuracy of five different classification approaches. In addition to accuracy, sensitivity, specificity, positive predictive value and negative predictive value were also compared. While these studies identified a suitable classification method for predicting the quality of wine with high accuracy, their results did not show the strength of correlation between physicochemical properties and the quality of wine. Yet, from the perspective of a wine manufacturer, understanding such a correlation is of great importance as this will allow them to make corrective measures that result in higher quality of the wine being produced. Unlike previous studies, this study focuses on discovering the correlation between sensorial and physicochemical properties of wine in the form of association rules. Experiments were conducted to evaluate the strength of the correlation in terms of rule confidence. Though the focus of this study is on association, rather than prediction, the rules generated can also be used for prediction purposes. Further, our study takes into account the computation cost for the wine industry. If the amount of association rules increases, it is more difficult for wine manufactuers to identify critical rules and the algorithm is also computationally expensive. In view of this, our study provides a solution with the use of constraints embedded in the algorithm so as to eliminate irrelevant rules. Table 8 shows the comparison between existing studies and this study.

Table 6. Summary of rules existing in all testing sets

	$\mu = 50\%$		$\mu = 60\%$		$\mu = 70\%$		$\mu = 80\%$		$\mu = 90\%$	
	No. of Rules	%	No. of Rules	%	No. of Rules	%	No. of Rules	%	No. of Rules	%
All common rules	982	100.00%	1018	100.00%	840	100.00%	684	100%	618	100.00%
Avg. $s>=1.5\%$	642	65.38%	638	62.67%	642	76.43%	502	73.39%	464	75.08%
Avg. $s>=2\%$	342	34.83%	336	33.01%	314	37.38%	294	42.98%	294	47.57%
Avg. $s>=3\%$	122	12.42%	118	11.59%	124	14.76%	116	16.96%	104	16.83%
Avg. $c>=55\%$	856	87.17%	902	88.61%	778	92.62%	630	92.11%	592	95.79%
Avg. $c>=60\%$	446	45.42%	498	48.92%	452	53.81%	432	63.16%	492	79.61%
Avg. $c>=70\%$	60	6.11%	76	7.47%	56	6.67%	86	12.57%	172	27.83%
Avg. $c>=80\%$	0	0.00%	0	0.00%	0	0.00%	4	0.58%	36	5.83%
Avg. $l>=1.5$	546	55.60%	582	57.17%	468	55.71%	350	51.17%	444	71.84%
Avg. $l>=2$	340	34.62%	340	33.40%	264	31.43%	280	40.94%	346	55.99%
Avg. $l>=2.5$	14	1.43%	10	0.98%	16	1.90%	36	5.26%	88	14.24%
No. of 2-itemsets	0	0.00%	1	0.10%	2	0.24%	1	0.15%	1	0.16%
No. of 3-itemsets	64	6.52%	63	6.19%	60	7.14%	54	7.89%	30	4.85%
No. of 4-itemsets	302	30.75%	298	29.27%	277	32.98%	220	32.16%	181	29.29%
No. of 5-itemsets	394	40.12%	404	39.69%	344	40.95%	269	39.33%	255	41.26%
No. of 6-itemsets	187	19.04%	206	20.24%	141	16.79%	121	17.69%	126	20.39%
No. of 7-itemsets	33	3.36%	42	4.13%	16	1.90%	19	2.78%	24	3.88%
No. of 8-itemsets	2	0.20%	4	0.39%	0	0.00%	0	0.00%	1	0.16%

5.2. Contributions

5.2.1. Conducting Association Rule Mining to Improve Wine Knowledge

In response to the challenges faced by the global wine makers (Gilinsky et al., 2016), such as the quality control during wine production (Niimi et al., 2017) for the most appropriate sensorial properties tailored to specific market segments (Biasoto et al., 2014), this paper proposed an RBQAS with the aim to fill the existing gap in wine production quality control which lacks of scientific approaches to understanding the relationships between production parameters and the resultant quality of the wine. Specifically, the proposed algorithm supports both descriptive and predictive analytics. For descriptive analytics, a set of rules is generated to discover the relationships among wine-related variables, such as the sensorial and physicochemical attributes of wine. For predictive analytics, the rules could be used to predict the quality of the wine based on the physicochemical properties. Both the descriptive and predictive analytics provide wine manufacturers decision support in determining the appropriate chemical parameters for better wine quality.

5.2.2. Using Search Constraints to Eliminate Irrelevant Rules for the Ease of Decision Making

One of the contributions of the proposed RBQAS is the feasibility and practicality of decision making. The constraints introduced in the RBQAS allow effective rule reduction, while it is impractical to make decisions based on a large amount of rules. It is especially important for real-time monitoring during wine production, where a large amount of data is generated. Without a systematic approach for rule filtering, the amount of rules could be massive in which a large proportion could be trivial or meaningless. From the perspective of wine manufacturers, irrelevant rules should be discarded, with attentions only paid to the most important rules for decision making. Besides, the outputs of

Table 7. Examples of valid rules extracted from training sets when μ = 60%

Rule	Avg. s	Avg. c	Avg. l
If{0.056<=CL<0.117}&{0.26<=VA<0.35}&{3.13<=pH<3.22}&{40<=FSO2<54}, then{Quality=Average}.	26.2	75.049	2.5019
If{0.056<=CL<0.117}&{0.35<=VA<0.51}&{CA>=0.46}, then{Quality=Average}.	27.2	76.784	2.5612
If{0.44<=SO4<0.59}&{20<=FSO2<29}&{CL<0.04}&{RS<3.15}, then{Quality=Satisfactory}.	24	77.502	1.7328

Table 8. Comparing existing work with this study

	Ribeiro et al. (2009)	Appalasamy et al., (2012)	Aich et al. (2018)	This Study
Analytics type	Predictive	Predictive	Predictive	Descriptive and predictive
Tools	Decision trees, Artificial Neural Networks, linear regression	Decision tree, Naïve Bayes	Recursive partitioning decision tree (PART), C4.5, PART, Bagging classification and Regression tree, Random Forest, Boosted C5.0	Association rule mining
Performance measure metrics	Accuracy	Accuracy	Accuracy, sensitivity, specificity, positive predictive value and negative predictive value	Support, confidence, lift
Correlation between chemical parameters and wine quality	û	û	û	ü
Computational cost consideration	û	ü (by feature selection)	ü (by feature selection)	ü(by eliminating irrelevant rules)

association rule mining, which are in the forms of 'if-then' rules are easy for users without technical or medical background, to interpret and to deploy quality control strategies based on the results.

5.2.3. Obtaining Rules of High Generalisation Potential for Global Applications

The proposed RBQAS allows the elimination of decision rules with poor generalization potential for wine quality management. Experiments conducted on a real data set illustrate the elimination of unreliable rules with validation on independent test sets. The algorithm performs several train and test iterations to achieve basic cross-validation and to reduce rules of poor generalization potential. The significance of the rules is evaluated in terms of support, confidence, and lift. From the methodological perspective, the RBQAS can discover a set of well-trained rules, representing the best practice of wine quality management. Wine manufacturers monitor and control the wine production conditions based on their experience to ensure the desirable resultant quality of wine. The knowledge discovered by the RBQAS thus supplements for wine manufacturers expertise and facilitates the wider applications of the set of rules.

5.3. Practical Implications

There are two key implications for practice. First, our results indicate that the different combinations of physicochemical parameter settings influence the sensorial property of the wine, which in turn affect the resultant quality of the wine. Therefore, there is a need for the wine industry to real-time monitor and control the parameters in wine production. Second, based on the rules obtained, it is found that the quality of white wine is highly associated with chlorides, volatide acidity, free sulfur dioxide, pH value, citric acid, sulphates and residual sugar. Hence, we suggest that white wine manufacturers should conduct a wider range of tests particularly towards these physicochemical parameters as white wine quality sesms to be affected by such substances.

6. CONCLUSION

This paper presents a rule-based quality analytics system for wine production with a constrained-rule mining algorithm embedded. In this proposed design, the number of rules obtained by the standard Apriori algorithm is much reduced which used to be one of the key challenges of advanced data technologies adoption. The sizes of the itemsets are constrained in the first phase of the algorithm while the appearances of items in the antecedent and consequent parts of the rules are constrained in the second phase on the algorithm. Moreover, the proposed algorithm combines search constraints and validation procedures so as to achieve cross-validation and reduce the number of rules with poor generalization potential. This paper serves as a good reference not only for the researchers in the field but also for the industrial partitiners in the global wine making business, with an attempt to improving the quality system with less hectic.

6.1. Limitations

There are three limitations of this paper. First, the focus of the paper is on the algorithm rather than the deployment of IoT which depends on the specific technical requirement of each project. Though the technical implementation of IoT is not discussed in the paper, the implementation issues have an impact on the efficiency and effectiveness of the system. Second, the dataset used for this study was collected for other purposes in a previous study by Cortez et al. (2009). We have no access to the original data set and cannot verify the quality of the data before applying the proposed algorithm. Nevertheless, the quality of the rules obtained can be affected by the quality of the data set we used. Thirdly, the data set used for this study is confined to white wine samples. The knowledge discovered may not be applicable to other types of wine. For instance, red wine is less sentistive to changes in physicochemistry properties as compared to white wine (Appalasamy et al., 2012). Attributes that influence the quality of white wine may not have the same effect on red wine quality. Thus, our results are sensitive to the types of wine.

6.2. Future Research Directions

Future research directions are threefold. Firstly, the constrained-rule mining algorithm can be integrated with other data mining techniques, such as fuzzy logic, to eliminate the drawback of using association rule mining (i.e. limitation to Boolean attributes, Alatas et al., 2008). Currently in the global winemaking industry, wines are described mainly according to the monthfeel wheels developed by Gawel et al. (2000) and Pickering and Demiglio (2008), for red and white wine respectively. Considering the attributes in winemaking generally involve numeric and linguistic descriptions, it could be useful to further introduce the fuzzy set concepts into the RBQAS, for exploring the relationships between sensorial and physicochemical properties of wine, say, hybridizing association rule mining with fuzzy set concepts. Secondly, the research can also be extended to the deployment of a sensor network in production and the enhancement of the system adaptability. The system can be implemented in the global wine industry for a pilot run and the IoT technologies can be deployed according to

the technical requirements for data acquisition. Thirdly, since the data used for the evaluation of the algorithm is confined to white wine samples, it will be interesting to apply the algorithm on other types of wine data, such as red wine data, and compare the rules. This will provide additional insights for the wine manufacturers on the optimization of production conditions for different types of wine.

FUNDING AGENCY

ACKNOWLEDGMENT

REFERENCES

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of 20th international conference on very large databases*, 487–489.

Aich, S., Al-Absi, A. A., Hui, K. L., Lee, J. T., & Sain, M. (2018). A classification approach with different feature sets to predict the quality of different types of wine using machine learning techniques. *2018 20th International Conference on Advanced Communication Technology (ICACT)*, 139-143.

Alatas, B., Akin, E., & Karci, A. (2008). MODENAR: Multi-objective differential evolution algorithm for mining numeric association rules. *Applied Soft Computing*, *18*(2), 646–656. doi:10.1016/j.asoc.2007.05.003

Altuntas, S., & Selim, H. (2012). Facility layout using weighted association rule-based data mining algorithms: Evaluation with simulation. *Expert Systems with Applications*, *39*(1), 3–13. doi:10.1016/j.eswa.2011.06.045

Appalasamy, P., Mustapha, A., Rizal, N. D., Johari, F., & Mansor, A. F. (2012). Classification-based Data Mining Approach for Quality Control in Wine Production. *Journal of Applied Sciences (Faisalabad)*, *12*(6), 598–601. doi:10.3923/jas.2012.598.601

Bastian, S. E. P., Collins, C., & Johnson, T. E. (2010). Understanding consumer preferences for Shiraz wine and Cheddar cheese pairings. *Food Quality and Preference*, *21*(7), 668–678. doi:10.1016/j.foodqual.2010.02.002

Biasoto, A. C. T., Netto, F. M., Marques, E. J. N., & Da Silva, M. A. A. P. (2014). Acceptability and preference driv4ers of red wines produced from during production has important consequences to consumer's acceptance and preferences of wine *Vitis labrusca* and hybrid grapes. *Food Research International*, *62*, 456–466. doi:10.1016/j.foodres.2014.03.052

Chang, C. P., Chu, C. P., & Yeh, Y. F. (2009). Integrating in-process software defect prediction with association mining to discover defect pattern. *Information and Software Technology*, *51*(2), 375–384. doi:10.1016/j.infsof.2008.04.008

Chougule, R., Rajpathak, D., & Bandyopadhyay, P. (2011). An integrated framework for effective service and repair in the automotive domain: An application of association mining and case-based-reasoning. *Computers in Industry*, *62*(7), 742–754. doi:10.1016/j.compind.2011.05.007

Coenen, F., Leng, P., & Ahmed, S. (2004). Data Structure for Association Rule Mining: T-Trees and P-Trees. *IEEE Transactions on Knowledge and Data Engineering*, *16*(6), 774–778. doi:10.1109/TKDE.2004.8

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, *47*(4), 547–533. doi:10.1016/j.dss.2009.05.016

Demiriz, A., Ertek, G., Atan, T., & Kula, U. (2011). Re-mining item associations: Methodology and a case study in apparel retailing. *Decision Support Systems*, *52*(1), 284–293. doi:10.1016/j.dss.2011.08.004

Gawel, R., Oberholster, A., & Francis, I. L. (2000). A 'Mouth-feel Wheel': Terminology for communicating the mouth-feel characteristics of red wine. *Australian Journal of Grape and Wine Research*, *6*(3), 203–207. doi:10.1111/j.1755-0238.2000.tb00180.x

Gilinsky, A. Jr, Newton, S. K., & Vega, R. F. (2016). Sustainability in the global wine industry: Concepts and cases. *Agriculture and Agricultural Science Procedia*, *8*, 37–49. doi:10.1016/j.aaspro.2016.02.006

Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, *29*(7), 1645–1660. doi:10.1016/j.future.2013.01.010

Hand, H., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT Press.

Jabnoun, N. (2002). Control processes for total quality management and quality assurance. *Work Study*, *51*(4), 82–190. doi:10.1108/00438020210430733

Kar, A. K., Joseph, N., Ilavarasan, P. V., & Ganesh, S. (2017, April). Review of Discussions on Internet of Things IoT: Insights from Twitter Analytics. *Journal of Global Information Management*, *25*(2), 38–51. doi:10.4018/JGIM.2017040103

Khatwani, G., & Srivastava, P. R. (2017). An Optimization Model for Mapping Organization and Consumer Preferences for Internet Information Channels. *Journal of Global Information Management*, *25*(2), 28. doi:10.4018/JGIM.2017040106

Kit, N. C., Ho, W. C., Leung, Y. K., Hung, I. W., & Tommy, C. (2018). A semantic similarity analysis of Internet of Things. *Enterprise Information Systems*, *12*(7), 820–855. doi:10.1080/17517575.2018.1464666

Köksal, G., Batmaz, İ., & Testik, M. C. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications*, *38*(10), 13448–13467. doi:10.1016/j.eswa.2011.04.063

Lee, C. K. H., Choy, K. L., Ho, G. T. S., Chin, K. S., Law, K. M. Y., & Tse, Y. K. (2013). A hybrid OLAP-association rule mining based quality management system for extracting defect patterns in the garment industry. *Expert Systems with Applications*, *40*(7), 2435–2446. doi:10.1016/j.eswa.2012.10.057

Lee, S., Park, J., & Kang, K. (2015). Assessing wine quality using a decision tree. *Systems Engineering (ISSE) 2015 IEEE International Symposium on*, 176-178.

Lim, A. H. L., Lee, C. S., & Raman, M. (2012). Hybrid genetic algorithm and association rules for mining workflow best practices. *Expert Systems with Applications*, *39*(12), 10544–10551. doi:10.1016/j.eswa.2012.02.183

Martínez-de-Pisón, F. J., Sanz, A., Martínez-de-Pisón, E., Jiménez, E., & Conti, D. (2012). Mining association rules from time-series to explain failures in a hot-dip galvanizing steel line. *Computers & Industrial Engineering*, *63*(1), 22–36. doi:10.1016/j.cie.2012.01.013

Ng, R., Lakshmanan, L. V. S., Han, J., & Pang, A. (1998). Exploratory Mining and Pruning Optimizations of Constrained Associations Rules. *Proceeding ACM SIGMOD Conference*, 13-24. doi:10.1145/276304.276307

Niimi, J., Danner, L., Li, L., Bossan, H., & Bastian, S. E. P. (2017). Wine consumers' subjective responses to wine mouthfeel and understanding of wine body. *Food Research International*, *99*, 115–122. doi:10.1016/j.foodres.2017.05.015 PMID:28784467

Ordonez, C. (2006). Association Rule Discovery With the Train and Test Approaches for Heart Disease Prediction. *IEEE Transactions on Information Technology in Biomedicine*, *10*(2), 334–343. doi:10.1109/TITB.2006.864475 PMID:16617622

Ouvrard, S., Jasimuddin, S. M., & Spiga, A. (2020). Does Sustainability push to reshape business model? Evidence from the European wine industry. *Sustainability*, *12*(6), 2561. doi:10.3390/su12062561

Pickering, G. J., & Demiglio, P. (2008). The White Wine Mouthfeel Wheel: A Lexicon for Describing the Oral Sensations Elicited by White Wine. *Journal of Wine Research*, *19*(1), 51–67. doi:10.1080/09571260802164038

Ribeiro, J., Neves, J., Sanchez, J., Delgado, M., Machado, J., & Novais, P. (2009). Wine vinification prediction using data mining tools. *Conference Proceedings Computing and Computational Intelligence*, 78-85.

Song, M., & Rajasekaran, S. (2006). A Transaction Mapping Algorithm for Frequent Itemsets Mining. *IEEE Transactions on Knowledge and Data Engineering*, *18*(4), 472–481. doi:10.1109/TKDE.2006.1599386

Srikant, R., Vu, Q., & Agrawal, R. (1997). Mining association rules with item constraints. *Proceedings KDD*, 67-73.

Tsang, Y., Choy, K., Wu, C., Ho, G., Lam, H., & Koo, P. (2017). An IoT-based cargo monitoring system for enhancing operational effectiveness under a cold chain environment. *International Journal of Engineering Business Management*, *9*. Advance online publication. doi:10.1177/1847979017749063

Ur-Rahman, N., & Harding, J. A. (2012). Textual data mining for industrial knowledge management and text classification: A business oriented approach. *Expert Systems with Applications*, *39*(5), 4729–4739. doi:10.1016/j.eswa.2011.09.124

Wang, K., He, Y., & Han, J. (2003). Pushing Support Constraints Into Association Rules Mining. *IEEE Transactions on Knowledge and Data Engineering*, *15*(3), 642–658. doi:10.1109/TKDE.2003.1198396

Xu, L. D., He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, *10*(4), 2233–2243. doi:10.1109/TII.2014.2300753

Carmen Kar Hang Lee is a Lecturer in the School of Business at Singapore University of Social Sciences (SUSS). Prior to joining SUSS, she was a Research Officer and an Honorary Assistant Professor in the Department of Industrial and Manufacturing Systems Engineering at the University of Hong Kong. She is active in research areas related to data analytics, operations management, logistics warehousing, supply network design and social big data. She was a visiting research scholar in the York Management School at the University of York in 2018. Her work has been published in journals such as International Journal of Production Economics, International Journal of Production Research, Engineering Applications of Artificial Intelligence, Expert Systems with Applications, Information Technology & People, Industrial Management & Data Systems among others. She serves as an Associate Editor for International Journal of Engineering Business Management.

Kris M. Y. Law is currently an Associate Professor at the School of Engineering, Deakin University, Australia. Prior to her joining Deakin University, she was a Lecturer at the Department of Industrial and Systems Engineering, Hong Kong Polytechnic University. She currently also holds a Docentship (Adjunct Professorship) in the Department of Industrial Engineering and Management, Oulu University in Finland. She undertook a post-doctoral research scholarship and was a Visiting Researcher at the Graduate Institute of Industrial Engineering, National Taiwan University in 2009–2011. Her expertise lies in organizational learning and development, technology and innovation management, technology-based entrepreneurship, project management and engineering education. She has published in journals and edited several books in the areas of performance management, organizational learning, technology and knowledge management.

Andrew W. H. Ip has more than 30 years of experience in teaching, research, education, industry and consulting. He received his PhD from Loughborough University (U.K.), MBA from Brunel University (U.K.), MSc in Industrial Engineering from Cranfield University (U.K.), and LLB (Hons) from the University of Wolverhampton (U.K.). He is now Professor Emeritus Mechanical Engineering of University of Saskatchewan, and Principal Research Fellow in the Department of Industrial and Systems Engineering of The Hong Kong Polytechnic University. He is also visiting professor of various universities in Mainland China, and Honorary Industrial Fellow of the University of Warwick, Warwick Manufacturing Group, UK. Prof. Ip has published nearly 250 papers with over 150 papers in SCI indexed journals. He is also the Editor-in-Chief of Enterprise Information Systems (SCI Indexed), and the Editor-in-Chief and Founder of International Journal of Engineering Business Management (ESCI and SCOPUS Indexed). He is a senior member of the Institute of Electrical and Electronics Engineers (IEEE), and member of the Hong Kong Institution of Engineers (HKIE).