Transactions on Machine
Learning and Data Mining
Vol. 10, No. 1 (2017) 3-24

ISSN: 1865-6781 ,
ISBN: 978-3-942952-47-7

ibai Publishing
www.ibai-publishing.org

A Wineinformatics Study for White-box Classification Algorithms to Understand and Evaluate Wine Judges

Bernard Chen[1], Hai Le[1], Travis Atkison[2], Dongsheng Che[3]

[1]Department of Computer Science University of Central Arkansas
Conway, AR, 72034, USA
[2]Department of Computer Science, University of Alabama,
Tuscaloosa, AL, 35487 USA
[3]Department of Computer Science East Stroudsburg University
East Stroudsburg, PA, 18301, USA
bchen@uca.edu

Abstract. Wineinformatics is a new data science research domain that utilizes wine as the domain knowledge. Wines are usually evaluated by wine judges who give scores to the wines they review. This paper proposes to use white box classification algorithms to understand why the wine judges score a wine as 90+ or 90-. Several white box classification algorithms with improved components are applied to wine sensory data derived from professional wine reviews. Each algorithm is able to tell how the judges make their decision. The extracted information is also useful to wine producers, distributors, and consumers. The dataset includes 1000 wines with 500 scored as 90+ points (positive class) and 500 scored as 90- points (negative class). Decision Tree, Association Classification, k-NN, Naïve Bayes and SVM are applied to the data and compared. The higher the accuracy retrieved from the algorithm, the more suitable it is for understanding the wine judges. The best white-box classification algorithm prediction accuracy we produced under 5-fold cross validation was 85.7% using Naïve Bayes algorithm with Laplace. The result indicates that the Naïve Bayes algorithm with Laplace might be the best white-box classification algorithm to understand wine judges. The SVM, a typical black-box classification algorithm, achieves 88% accuracy. Sensitivity and specificity are also evaluated in selected algorithms. To the best of our knowledge, it is the first time that the classification algorithms are applied and compared in wine sensory reviews.

Keywords: Wineinformatics, White-box Classification, Decision Tree, Association Classification, Naïve Bayes, k-Nearest Neighbors, SVM.

1 Introduction

Data mining is often set in the broader context of knowledge discovery in databases, or KDD. This term originated in the artificial intelligence (AI) research field. The KDD process involves several stages: selecting the target data, preprocessing the data, transforming it if necessary, performing data mining to extract patterns and relationships, and then interpreting and assessing the discovered structures. It is most useful in an exploratory analysis scenario in which there are no predetermined notions about what will constitute an "interesting" outcome. Best results are achieved by balancing the knowledge of human experts in describing problems and goals with the search capabilities of computers [1].

The earliest evidence of **wine** making was found in China in 7000 BCE based on fermented honey, rice, and fruit. Since then, with the development of society and the rise in standard of living, the qualities and varieties of wines are increasing year by year. With the development of society, and as quality of life rises, the qualities and varieties of wines are increasing year by year. According to OIV (International Organization of Wine and Vine) [2] estimates, the 2016 world wine production was estimated at 259 million hl (1 hl = 100,000 ml) [3]. Although it is considered a large number, the 2016 production is considered the lowest production for the past 20 years as a consequence of climatic events. In accordance with this information, wine is one of the most widely consumed beverages in the world and has very obvious commercial value as well as social importance. Therefore, the evaluation of the quality of wine plays a very important role for both manufacture and sale [4]. An established approach to investigate which aspects have significant effects on willingness to pay for food products is to focus on objective characteristics (such as price, brand, and appearance), consumer demographics (such as age, income and education level), and frequency of consumption. Sensory properties such as taste, aroma, texture, and flavor are typically not included. However, sensory qualities are often the major factors that affect consumers' perception of a product. Therefore, it is necessary to include them in assessing consumer's preference [5].

To better analyze wines, reputable wine reviewers from professional wine magazines, such as Wine Spectator, Wine Advocate, Wine Enthusiast, and Decanter, use human language to describe them in great detail. Here is an example:

Kosta Browne Pinot Noir Sonoma Coast 2009 *95pts*

Ripe and deeply flavored, concentrated and well-structured, this full-bodied red offers a complex mix of black cherry, wild berry and raspberry fruit that's pure and persistent, ending with a pebbly note and firm tannins. Drink now through 2018. 5,818 cases made. [16]

These reviews are based on the sensory attributes conveyed by a wine, and they cover a broad range of detail; acidity, flavor, color, and smell are just a few examples of the attributes that wine reviewers take into consideration to describe a wine. All of the professional wine magazines have large databases to store their historical wine reviews; for example, Wine Spectator contains more than 300,000 wine reviews available for paid members. Although hundreds of thousands of different wine reviews are stored in each magazine's database, very limited amount of data mining research has been applied in this interesting field.

This paper is interested in understanding how wine experts review the wines through white-box classification algorithms; also, the developed models can then be used to evaluate the wine judges. According to American Association of Wine Economics, "*Who is a reliable wine judge? How can we aggregate the will of a tasting panel? Do wine judges agree with each other? Are wine judges consistent? What is the best wine in the flight?*" are typical questions that beg for formal statistical answers [6]. Some researchers work on this problem by looking into ranking, rating, and judging of the wine through traditional statistical methods [6 - 9]. This paper is a continuing work on a new data science research area named Wineinformatics, which uses the understanding of wine to serve as the domain knowledge [23]. We convert the wine savory reviews through the computational wine wheel, and then we apply different data mining white-box classification algorithms to the same dataset. Our goal is to find the best white box classification algorithm to understand and evaluate the consistency of wine judges.

Different white-box classification algorithms can provide distinct useful information. Classification consists of predicting a certain outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute [12]. Decision tree uses a predictive model to determine consequences. The application of boosting procedures to decision tree algorithms has been shown to produce very accurate classifiers [11]. Association rules were initially made popular by market-basket analysis [24]. The algorithm was developed in order to discover the connection between items in a purchase for large transactional databases. While we associate the wine attributes with the wine quality, we can form the association classification in Wineinformatics. k-nearest neighbor (k-NN) focuses on how each wine is similar to each other, divides all similar wines into clusters, and predicts the accuracy of the data [18]. Finally, Naïve Bayes is a statistical classifier to predict class membership probabilities, such as the probability that a given tuple belongs to a particular class [10, 20]. Some of the variations of Naïve Bayes models are used for text retrieval and classification, focusing on the distributional assumptions made about word occurrences in documents [12].

These white-box classification algorithms will combine with the real data to classify the wine into different categories. Although these classification algorithms are considered textbook algorithms, it is the first time that they are applied in Wineinformatics to the best of our knowledge. We also provide our insight of how to use white-box classification models to benefit wine makers, distributors, and consumers. Last but not least, we compare the results generated from SVM, which is a typical black-box classification algorithm, with all white box classification algorithms for the purpose of the benchmark comparison.

The framework of this paper will be laid out as follows: Section 2 will introduce the data of wine in detail; Section 3 will describe how different classification methods—Decision Tree, Association Classification, k-NN, and Naïve Bayes, work on the data; Section 4 will demonstrate the result and the accuracy between different classification algorithms; finally, we will cover the conclusion and future works in the last section.

2 Dataset for the Experiments

2.1 Wine Sensory Data

The evaluation of wine can be categorized into two major methods. The first is of an analytical instrumental sequence using spectroscopic and chromatographic methods where the wine is analyzed for its chemical compounds [15]. The second is of sensory qualifications which is a professional wine reviewer perceives via organoleptic properties – these being the aspects as experienced by the senses of taste, sight, and smell [13]. Figure 1 provides an example for a wine review by both perspectives.

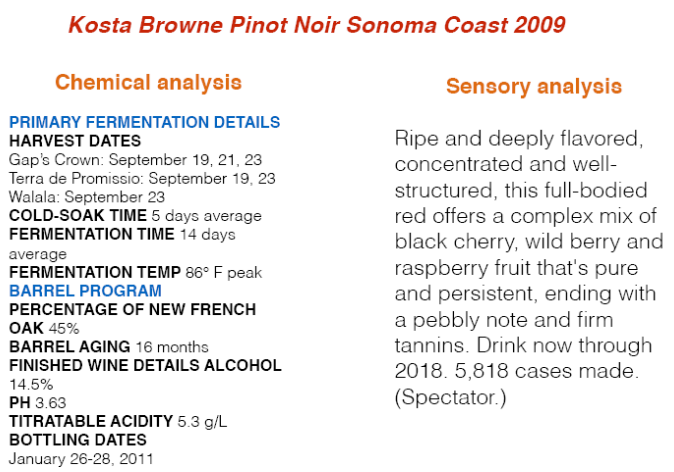

Fig. 1. The review of the Kosta Browne Pinot Noir Sonoma Coast 2009 (scores 95 pts) on both chemical and sensory analysis.

Physicochemical laboratory tests [25, 26] routinely used to characterize wine include determination of density, alcohol or pH values, while sensory tests rely mainly on human experts [26]. Most of the existing data mining/data science research related to wine [26 - 28] focuses on the physicochemical laboratory tests data. However, in wine economics point of view, sensory analysis is much more interesting to consumers and industrial perspective than chemical analysis since they describe aesthetics, pleasure, complexity, color, appearance, odor, aroma, bouquet, tartness, and the interactions with the senses of these characteristics [29] of the wine.

Wine sensory analysis involves tasting a wine and being able to accurately describe every component that makes it up. Not only does this include flavors and aromas, but characteristics such as acidity, tannin, weight, finish, and structure. Within each of those categories, there are multitudes of possible attributes or forms that each can take. What makes the wine tasting process so special is the ability for two people to simultaneously view the same wine while being able to share and detect all the same attributes.

2.2 Wine Spectator

Wine reviews are made of the most sensitive and critical sensory evaluation techniques, which have little room for error, and quality control is critical [16]. Although there are many different wine expert reviews, such as *Wine Advocate*, *Decanters Magazine*, or *eRobertParker*, the data in this paper is derived from the Wine Spectator magazine's wine sensory data. We used the Wine Spectator data source primarily for its impact on the wine culture due to its extensive wine reviews, ratings and general consistency not to logomachy in wine reviews. Wine Spectator publishes 15 issues per year, and each issue contains between 400 and 1000 wine reviews [15]. The reviews are direct and specific to the sensory perception of the wine. The wine tests are blind tests in controlled environments, and reviewers are only aware of the type of wine and vintage. Reviews consist of the 50 – 100 point scale in which wine professionals grade each wine against other wines in its same category for overall quality. Reviews also consist of the sensory attributes of each individual wine. These sensory attributes are where we pull our dataset from. Wine Spectator tasters review wines on the following 100-point scale as showed in Table I.

Table I. 100-point scale of Wine Spectator.

Score	Classification	Description
95-100	Classic	a great wine
90-94	Outstanding	a wine of superior character and style
85-89	Very Good	a wine with special qualities
80-84	Good	a solid, well-made wine
75-79	Mediocre	a drinkable wine that may have minor flaws
50-74	Not recommended	Not recommended

An honor is given to a wine when it is scored above 90 points (a great wine or a wine of superior character and style). Consistency of evaluation is the key to maintaining the reputation of wine judges. The goal of this paper is to understanding how wine experts review the wines. To achieve this goal, we plan to use white-box classification algorithms to build models based on Wine Spectator's reviews. The performance of the models can be considered as the criteria to evaluate Wine Spectator as the wine judge; the more consistent the wine judge, the higher performance classification models can perform. Since there are many different classification algorithms available, this paper also tries to identify the best algorithm for understanding the wine judges and evaluating their consistency in the Wineinformatics application domain.

2.3 Dataset for Experiments

Hundreds of thousands of professional wine reviews are published in human language format each year. It is impossible to read and process all the reviews manually. As a result, we developed a natural language processing tool named the Computational Wine Wheel [23] to automatically extract key attributes from wine sensory reviews. The purpose of the Computational Wine Wheel is to not only capture all flavors but also feeling expressions as described in the experts' reviews. In our opinion, those key terms play important roles in our research as well. For example, if APPLE flavor appears in both 91 points and 82 points wines, words such as WELL-STRUCTURED, BEAUTIFUL, or AGE WELL might show the difference between them. Therefore, after the process of the computational wine wheel has been applied to the review in Figure 1, all the terms that are in bold will be extracted and considered characteristics of the wine.

Kosta Browne Pinot Noir Sonoma Coast 2009

Ripe *and* ***deeply flavored***, ***concentrated*** *and* ***well-structured***, *this* ***full-bodied*** *red offers a complex mix of* ***black cherry***, ***wild berry*** *and* ***raspberry*** *fruit that's* ***pure*** *and* ***persistent***, *ending with a* ***pebbly*** *note and* ***firm tannins***. *Drink now through 2018. 5,818 cases made.*

In this research, the dataset includes a multi-year span that consists of 1000 wine sensory reviews, including 500 wines scored 90+ and another 500 wines scored 90-. The reviews are scanned word by word through the computational wine wheel [14, 23]. If there is a match word in the review with the "specific name" in the Computational Wine Wheel, the "categorized name" attribute is assigned positive to the wine. For example, if a wine review has FRESH-CUT APPLE or RIPE APPLE or APPLE, these wine attributes are categorized into a single category APPLE. However, according the Computational Wine Wheel, GREEN APPLE is considered as GREEN APPLE, which is not in the APPLE category since the flavor is different. According to [23], the Computational Wine Wheel contains 304 normalized attributes. The dataset can be visualized as Table II. If a wine review for an individual wine contained an attribute, a 1 was listed in the column for that attribute for that wine to indicate ‘true’; otherwise a 0 was listed for ‘false’. Also, the wines were given a classification on the 100-point scale. If a wine scores equal or higher than 90 points, we consider it as a positive class; on the other hand, if a wine scores below 90 points, we consider it as a negative class. In Table II, the first 250 wines were in the [95-100] scores category, the next 250 wines were in the [90-94] scores category, the next 250 wines were in the [85-89] scores category and the last 250 were in the [80-84] scores category. By using the Kosta Browne Pinot Noir Sonoma Coast 2009 wine mentioned earlier as an example, the attributes extracted are: RIPE, CONCENTRATED, FULL-BODIED, BLACK CHERRY, WILD BERRY, RASPBERRY FRUIT, PURE, PERSISTENT, PEBBLY, and FIRM TANNINS. Plus, this wine is considered as a positive class.

Table II. A visualized representation of the wine dataset.

304 attributes

	ACCENTS	ACIDITY	AGE WELL	ALLURING	ALMOND	ANISE	...	...	WHITE FRUIT	WHITE PEACH	WHITE PEPPER	WILD BERRY	WONDERFUL	YOUNG	
CHÂTEAU DE BEAUCASTEL Châteauneuf-du-Pape White Vieilles Vignes	0	1	0	0	0	0	...	...	0	0	0	0	0	0	
CLOS DES PAPES Châteauneuf-du-Pape White	0	0	0	0	1	0	...	...	0	1	0	0	0	0	
CHÂTEAU DE BEAUCASTEL Châteauneuf-du-Pape White Vieilles Vignes	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
DOMAINE DE BEAURENARD Châteauneuf-du-Pape White Boisrenard	0	0	0	0	0	0	...	...	0	0	0	0	0	0	250: [100-95]
C.H. BERRES Riesling Trockenbeerenauslese Mosel Ürziger Würzgarten	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
VIA WINES Pinot Noir Maule Valley Chilensis Reserva	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
VIA WINES Chardonnay Maule Valley Chilensis Reserva	0	0	0	0	0	0	...	...	0	0	0	0	0	0	250: [94-90]
VIA WINES Cabernet Franc-Carmenère Maule Valley Oveja Negra Reserva	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
VIA WINES Pais Maule Valley Chilcas Single Vineyard	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
VIA WINES Merlot Maule Valley Chilensis Reserva	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
VIA WINES Carmenère Maule Valley Chilensis Reserva	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
VIA WINES Carmenère-Merlot Maule Valley Oveja Negra Reserva	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
VIA WINES Chardonnay Casablanca Valley Chilcas Single Vineyard	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
VIGNERONS LAUDUN & CHUSCLAN Côtes du Rhône White La Ferme de Gicon	0	0	0	0	0	0	...	...	0	0	0	0	0	0	250: [85-89]
GASPARE VINCI Alcamo White	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
WATTLE CREEK Viognier Alexander Valley	0	1	0	0	0	0	...	...	0	0	0	0	0	0	
WESTWOOD FAMILY Benjamin's Cuvee Sierra De Montserrat Vineyard Placer County	0	0	0	0	0	0	...	...	0	0	0	0	0	0	250: [80-84]
WIENINGER Qualitätswein Trocken Wien Rosé de Pinot	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
WIENINGER Qualitätswein Trocken Wien Wiener Gemischter Satz	0	0	0	0	0	0	...	...	0	0	0	0	0	0	
WILDEKRANS Sauvignon Blanc Bot River Lot 1982	0	0	0	0	0	0	...	...	0	0	0	0	0	0	

3 Methods and Results

3.1 Decision Tree

Decision Tree induction is the learning of decision trees from class-labeled training tuples [17, 22]. The tree consists of nodes that form a rooted tree, meaning it is a directed tree with a node called “root” that has no incoming edges. All other nodes, called internal nodes, have exactly one incoming edge that denotes a test on an attribute, but it splits or branches to represent an outcome into two edges according to the input variable. Each leaf node holds a class label or an attribute. The Decision Tree algorithm is a tree that is constructed in a top-down recursive divide and conquer manner. In the beginning, all attributes are listed at the root. To determine which attribute is to become the root, we used a statistical measure called information gain. The attribute with the highest information gain is the root of the tree.

Table III. Example dataset to apply Decision Tree.

Name	CHERRY	APPLE	PURE	BERRY	Grade
Wine1	1	1	0	1	**90+**
Wine2	0	0	1	1	**90+**
Wine3	1	0	0	1	**90+**
Wine4	0	1	1	1	**90-**
Wine5	1	0	0	0	**90-**
Wine6	0	1	0	1	**90-**

The dataset shown in Table III has 6 wines and 4 attributes. Among 6 wines, the first 3 are graded 90+, and the last 3 are graded 90-. Names on the first row represent wine attributes: A: CHERRY, B: APPLE, C: PURE, and D: BERRY. After we apply the decision tree algorithm to the dataset showed in Table III, the generated tree is shown in Figure 2. Since the dataset has 2 classes, the decision tree becomes a binary tree. Due to the fact that attribute D (BERRY) gets the highest gain information, it becomes the root of the tree. Next is attribute A (CHERRY), and then attribute B (APPLE).

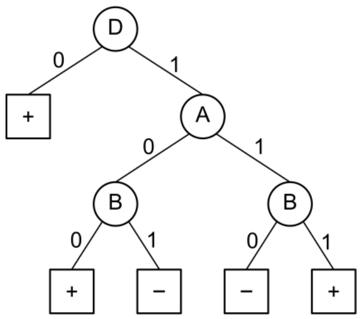

Fig. 2. The decision tree from Table III.

The decision tree can be used to predict the grade of a testing wine. For example, a testing wine has the following attribute: A(0), B(1), C(0), D(1). Since the root of the tree is attribute D, we check the value of attribute D of the Testing wine. Since it is 1, we follow path 1 and reach attribute A. We check the value of attribute A of the Testing wine; since it 0, we follow path 0 and reach attribute B. Again, we check the value of attribute B of the testing wine; since it is 1, we follow path 1 and reach the bottom of the tree. At this point, it stops and predicts that the testing wine has a 90- grade.

The benefit of the decision tree is that the mined information has high readability. Important attributes are displayed on top of the tree. The prediction results are based on the combination of the attributes. Wine makers can use this information to decide their fermentation method and tools (such as French barrel or American barrel) to avoid bad combinations and improve the quality of the wine.

After we apply the 5-fold cross validation to our 1000 wine dataset with the Decision Tree algorithm, the average accuracy just barely passes 50% (50.6%). Since the problem we are facing is a typical bi-class classification problem, the accuracy is just better than guessing "heads" or "tails" when we flip a coin. We noticed a significantly lower percentage of 90- wines that were predicted compared to the 90+ wines. This could be due to the fact that 90- wines do not have as many of the attributes listed as the 90+ wines, which would cause problems with classifying by an attribute. As mentioned above, depending on the dataset, some classification algorithms will generate high accuracy predictions, and some will not; thus, decision tree is not suitable for the wine dataset tested in this paper. As a result, it gives us motivation to try and test more classification models.

3.2 Association Rules

Association rules were initially made popular by market-basket analysis [24]. The algorithm was developed in order to discover the connection between items in a purchase for large transactional databases. While made popular initially for marketing strategies, the algorithm can be useful for finding many relational insights in data. The association rules algorithm generates the rules in a form of A=>B. The rule A=>B holds both support s, which is the probability that a transaction contains A∪B, and confidence c, which is the conditional probability that a transaction having A also contains B. The formula is given below:

$$Support(A => B) = \frac{|A \cup B|}{|T|}$$

$$Confidence(A => B) = \frac{|A \cup B|}{|A|}$$

Rules that pass both user-defined minimum support and minimum confidence thresholds are called **Strong Association Rules**. In this paper, we make use of the association rules technique in generating frequent item-sets in order to reveal the underlying patterns in wine profiles. Each wine review is considered as a transaction with the paired attributes acting as items of a transaction. As each review is processed, the attributes are recorded to build a collection of frequent wine descriptors. At this point in the algorithm, frequent item-sets would be used to find a correlation to a "label" in the transaction, thus finding when another item should be present based on the item-sets found. This so called "label" could be extra information other than wine attributes, such as wine grade, region, grape type, etc. We choose wine grade as the label in this section. The goal is to accurately predict a wine scored above 90 (classic and outstanding wines) or below 90 (good and very good wines) based on only the sensory review.

Once the rules are formed, for each wine that we try to predict, we try to match the rules that apply to the predicting wine. If we find an applicable rule, we can predict the wine's grade category (90+ or 90-). If no rules are applicable, we cannot predict the wine. A third scenario is that more than one rule are applicable; in this case, we use the highest confidence rule to perform the prediction.

Table IV. Example of wine score range prediction via association rules algorithm.

	BLUEBERRY	CHERRY	CHEWY TANNINS	BEAUTY	Score
Wine 1	0	1	1	1	95
Wine 2	0	0	0	1	85
Wine 3	0	1	0	0	88
Wine 4	0	1	0	1	91
Wine 5	1	1	0	1	?

We also provide the following example in Table IV to show how we apply association rules to predict the range of the wine score. In the example, we have five wines in total, four of them with known scores and we try to predict the score range (>90 or <=89) of Wine 5. Assume we define minimum support=50% and minimum confidence=80%. Based on wines 1-4, we can find one strong association rule (CHERRY and BEAUTY => >90) with support=2/4 and confidence=2/2. The rule indicates "if a wine has cherry and beauty in their review, it is a 90+ points wine." Therefore, since this rule is applicable to Wine 5, we then can predict it as a 90+ points wine.

In the association rules algorithm, users need to define the minimum support and confidence. Different user defined values will produce different results. In this experiment, the higher minimum support and confidence value, the more rules will be generated; thus, the more wines can possibly be predicted; however, the prediction accuracy may drop. Table V provides the 5-fold cross validation prediction accuracy and coverage results according to different minimum support and confidence values.

In Table V, it is clear to see that with the same minimum support value, the best prediction accuracy (85.25%) is generated by minimum support=2% and minimum confidence=90%; however, the coverage is the lowest (50.90%); which means among 200 wines in the testing dataset, only 102 wines can be predicted. On the other side, the lowest prediction accuracy (72.25%) in the table is generated by minimum support=0.5% and minimum confidence=60% with 98.40% coverage. The results show a similar trend in each support column. The more restrictive requirement of associations results in a higher accuracy and lower coverage due to the decreased amount of rules being generated.

Table V. Experimental Results of prediction accuracy and coverage for 5-fold cross validation based on different combinations of minimum support and confidence.

Confi-dence	*0.5% Support*		*1.0% Support*		*1.5% Support*		*2.0% Support*	
	Accuracy	*Coverage*	*Accuracy*	*Coverage*	*Accuracy*	*Coverage*	*Accuracy*	*Coverage*
60%	*72.28%*	*98.40%*	*72.86%*	*97.30%*	*72.47%*	*95.80%*	*73.19%*	*94.80%*
70%	*73.07%*	*94.60%*	*74.50%*	*91.90%*	*75.15%*	*88.30%*	*76.48%*	*85.40%*
80%	*74.52%*	*87.10%*	*76.10%*	*80.30%*	*76.94%*	*75.50%*	*78.22%*	*71.00%*
90%	*78.10%*	*73.50%*	*82.27%*	*61.90%*	*83.01%*	*56.40%*	*85.25%*	*50.90%*

For the association classification algorithm, there is a possibility that more than one rule is applicable to the predicting wine while we try to perform the association classification. In order to seek the opportunity to further improve the prediction results, we also implemented a "majority vote" approach to compare against the initial "highest confidence" approach: If more than one rule can be applied to the testing wine, we predict the wine's score category by the majority of rules (if there is a tie, we take rules confidence into consideration). This can be easily understood by the following two examples presented in Tables VI and VII:

Table VI. Example of TURKEY FLAT Butchers Block White Barossa Valley for wine classification.

TURKEY FLAT Butchers Block White Barossa Valley ***91pts*** *Fresh and inviting, this delivers a juicy mouthful of pear, fresh grape and subtle spice flavors, which persist on the generous finish. Marsanne, Viognier and Roussanne.* Attributes: PEAR,SPICE,FRESH,GENEROUS,JUICY,SUBTLE	
Applicable Rules:	Rule Confidence:
Generous => 90+	93.75%
Fresh, Pear => 90-	73.91%
Fresh, Juicy => 90+	72.73%
Fresh Spice => 90+	70.00%

Table VII. Example of BODEGAS FINCA NUEVA Rioja White Ferentado en Barrica for wine classification.

BODEGAS FINCA NUEVA Rioja White Fermentado en Barrica ***86pts***	
This silky white is fresh and lively, with lime and pineapple flavors brightening a core of pear and herb. The sparkly acidity will need food for balance. Attributes: LIME,PEAR,PINEAPPLE,HERBS,WHITE,ACIDITY,BALANCE,FRESH,LIVELY,SILKY	
Applicable Rules:	Rule Confidence:
Acidity, Balance => 90+	95.83%
Fresh, Lime => 90-	92.86%
Lime, Pear => 90-	83.33%
Fresh, Herbs => 90-	81.82%
Balance => 90+	80.82%

In both examples, if we just apply the “highest confidence” approach, both will be predicted as 90+ points wine. In this case, the second wine is predicted in the wrong category. However, if we use the “majority vote” mechanism, the first wine will be considered as 90+ since 3 out of 4 rules suggest this is a 90+ wine; the second wine will be considered as 90- since 3 out of 5 rules suggest so. In our experiment, the “majority vote” did slightly increase the prediction accuracy with the coverage. The comparison results are generated using the 1% support category and run for each confidence measure from above 60%. The findings are displayed in Table VIII below, note that coverage is not displayed as there is no change while only re-interpreting the association rules at the coverage listed in Table V.

Table VIII. Comparison of prediction accuracy generated by “Highest confidence” and “Majority Vote” methods at 1% minimum support setup.

Conf.	*Highest confidence’s*	*Majority Vote’s*
60%	72.86%	73.79%
70%	74.50%	75.89%
80%	76.10%	76.24%
90%	82.27%	82.58%

3.3 k-Nearest Neighbors (k-NN)

k-Nearest Neighbors (k-NN) is “a non-parametric method used for classification and regression.” In both cases, the input consists of the k closest training examples in the feature space [19]. The output of the algorithm is a class membership, and an object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k > 0) [18]. In other words, k-NN does not build any model. k values are chosen, and the algorithm calculates distances between instances and then predicts labels directly.

For our wine dataset, the prediction of a test wine is based on the majority label vote of its k "nearest" wines. In other words, the algorithm chooses k wines that are the most similar to the test wine, counts how many of them are "90+" and "90-", and then predicts the test wine label based on the majority vote. As mentioned in the previous section, our wine dataset is in binary format. For that reason, Jaccard's distance formula is used.

Jaccard's distance formula: $J = \dfrac{Q+R}{P+Q+R}$

Q is the number of positive attributes in Wine 1 but not in Wine 2; R is the number of positive attributes in Wine 2 but not it Wine 1; and P is the number of positive attributes in both Wine 1 and Wine 2. The smaller the value is, the more similar the two wines are. For example, the Jaccard's distance between Wines 1 and 2 in Table IV is 2/3 (Q:2, R:0, P:1).

We tested our k parameter from 1 to 21 with an interval of 1. Because the wine dataset has 2 class labels (90+ and 90-), we choose k being an odd number to prevent equal voting. Figure 3 shows the 5-fold cross validation results of k-NN for each fold and its average when k = [1, 21]. The highest accuracy is 74.9% (k = 19), which is much better than the Decision Tree result of 50.6%. Overall, the accuracy results of k-NN are similar to each other except when k = 1. Because k-NN predicted the result based on the majority vote, with k = 1, the algorithm will be completely based on the label of only one wine. As a result, it leads to bias when the algorithm does not consider more instances to vote. Therefore, k = 1 is an outlier.

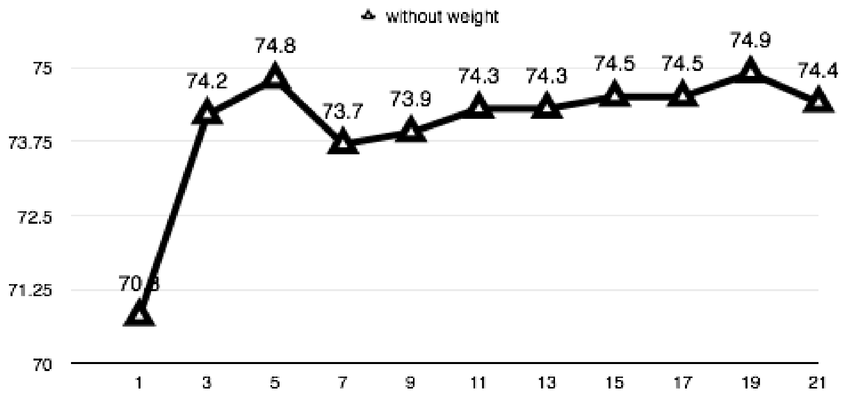

Fig. 3. The average accuracy of k from 1 to 21 (k-NN).

In this paper, we further study the weight contribution of different attributes in the k-NN algorithm. The Computational Wine Wheel [23] has one column about the attributes' category, where attributes are weighted differently (3, 2, and 1) based on their category. "1" is non-flavor descriptions (PURE, BEAUTY, WONDERFUL, etc.). "2" is the non-flavor wine characteristics (TANNINS, ACIDITY, BODY, etc.). "3" is the food and wine flavor characteristics (specific fruit, woods, flavors, etc.). Table IX gives an example of it.

Table IX. Simplified example for our wine dataset with weight.

Name	**LONG FINISH**	**APPLE**	**PURE**	**Grade**
Weight	2	3	1	
Wine1	0	1	1	90+
Wine2	1	0	0	90+
Wine3	1	1	1	90-
Wine4	0	1	0	90-
Test Wine	1	0	1	?

In Table IX, since we added the weight concept, the Jaccard's distance needs to be adjusted to include it:

$$J = \frac{weightQ + weightR}{weightQ + weightR + weightP}$$

where

weightQ = the sum of the number of positive attributes in Wine 1 but not in Wine 2 multiply the weight

weightR = the sum of the number of positive attributes in Wine 2 but not in Wine 1 multiply the weight

weightP = the sum of the number of positive attributes in both Wine 1 and Wine 2 multiply the weight

For example, the weighted Jaccard's distance between the testing wine and Wine 1 in Table XI is 5/6 ((2+3)/(2+3+1)). The distances between the testing wine and Wines 2, 3, 4 are 1/3((1+0)/(1+0+2)), 1/2((0+3)/(0+3+3)), 5/5((3+3)/(3+3+0)), respectively. For this specific k-NN example, if k=1, we will predict the testing wine belongs to 90+ since the closest wine is Wine 2. If k=3, we will predict the testing wine belongs to 90+ since the closest 3 wines are Wines 1(90+), 2(90+), and 3(90-).

Although different weights on three categories are provided in [23], the accuracy performance of k-NN may not follow the same logic. To prevent all the assumptions and biases, we switch the values of weights between attributes, and we create all possible combinations between them. Table X shows how it is done and Figure 4 depicts the 5-fold cross validation prediction accuracy results among all weight combinations.

Table X. All combinations of weight for different category attributes.

	Category 1 Attributes	Category 2 Attributes	Category 3 Attributes
original weight	1	2	3
combination 1	1	2	3
combination 2	3	2	1
combination 3	2	3	1
combination 4	2	1	3
combination 5	1	3	2
combination 6	3	1	2

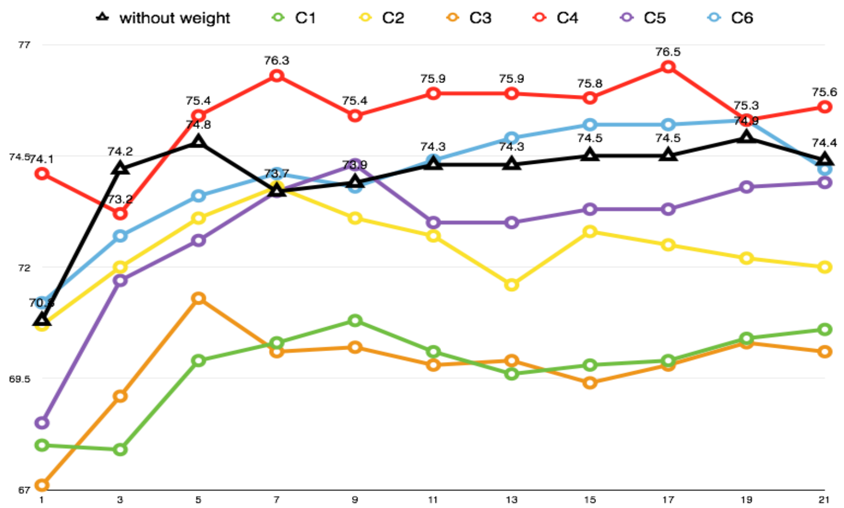

Fig. 4. The accuracy comparison chart of all weight combinations and without weight.

Among 6 combinations, combination 3 gives the lowest accuracy (67.1%) while combination 4 gives the highest result (76.5%). Compared to the highest accuracy without weight (74.9%), the original weight (combination 1) generates lower accuracy with the highest result being 70.6%. Combinations 4 and 6 are the two that perform better than the dataset without weight.

Based on the results of the experiment, combination 4, which generated the best accuracy among all, suggests that attributes with weight 3 are kept the same; but attributes that are weighted 1 are actually more important that those attributes that are weighted 2, so we need to switch them. Combination 6 follows the same routine, but it says the attributes weighted 1 are the most important. In both cases, even though there is a conflict between the original weight 1 and 3, all combinations agree that the attributes that are weighted 2 should be the least important attribute.

3.4 Naïve Bayes

A Naïve Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem [20] with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In other words, a Naïve Bayes classifier assumes that the presence (or absence) of an instance of a class is unrelated to the presence (or absence) of any other instance [20]. For example, a wine may be considered to be a "90+" wine if it has BLUE BERRY, APPLE, and LONG FINISH. Even if these attributes depend on each other or on other attributes, when a Naïve Bayes classifier generates the probability of the wine, it considers all of these attributes independently. As a result, depending on the precise nature of the probability model, Naïve Bayes classifiers can be trained very efficiently in a supervised learning setting.

In the Naïve Bayes algorithm, zero frequency happens when none of the training instances have the same value as the testing instances; therefore, the result will equal zero and ignore all the effects of other instances. There are several solutions to minimize the effect of zero frequency problems. We choose Laplace as the implementation in this paper. Laplace is a smoothing data technique; the purpose is manipulating the value of the data at the beginning, so Naïve Bayes classification will never have zero frequency problems. (Except when parameter k = 0). Bayes' theorem formula (1) will be modified to:

$$P(H \mid X) = \frac{P(X \mid H)P(H) + k}{P(X) + b}$$

where b is the number of instances in the dataset and k is the adjustable parameter. When applying the Add penalty and Laplace methods, we manipulate the value of k from 1 to 20. After we apply 5 fold cross validations, the results of both methods are shown in Figure 5.

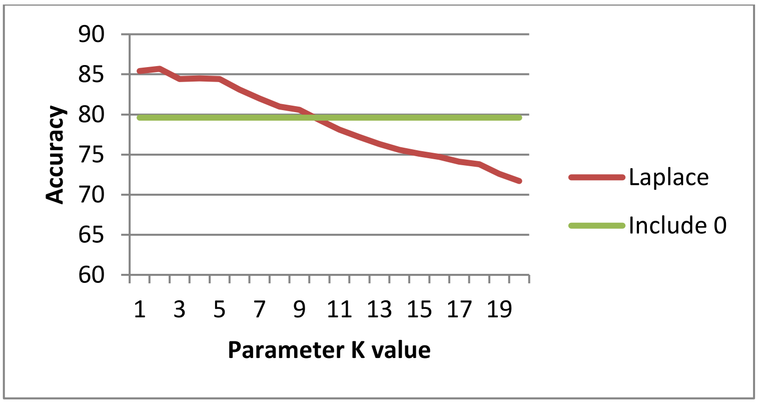

Fig. 5. The comparison between original Naïve Bayes (include 0 frequency) and Laplace.

Overall, Naïve Bayes generates great results. The accuracy is better than 80% in Laplace method with lower K values, which is quite high for a real dataset. However, Laplace shows that accuracy decreases when k increases. The best result in Laplace is 85.7% when K=2. For the original Naïve Bayes classification, since there is no k parameter in the formula, there is only one accuracy result = 79.6%. The satisfactory accuracy indicates that Wine Spectator does have a consistent review.

4 Prediction Accuracy Comparison with Published Results and SVM

White-box and black-box are two major categories of classification algorithms. To be able to analyze the prediction accuracy and draw out useful information from how the models react to the database, the classification models that we have covered so far are all white-box classification algorithms. Black-box classification algorithms, on the other hand, usually generate better results than white box testing, but we will not be able to explain how it obtains the conclusions. To complete our results comparison, Support Vector Machine (SVM), which is one of the most popular black box testing methods, is applied to our wine dataset for benchmark purpose.

During our SVM implementation, we made some minor improvements in scaling the dataset and choosing the best parameter, suggested by [21]. Scaling the dataset is to "avoid attributes in greater numeric ranges dominating those in smaller numeric ranges" and "avoid numerical difficulties during the calculation" [21]. As the paper suggests, we linearly scale each wine attribute to the range [-1, +1]. For the best parameter method, SVM used grid search to scan through the whole dataset and tried to pick the best C and Υ (C is penalty parameter, and Υ is kernel parameters). Table XI shows the results produced by SVM and Figure 6 gives the final comparison results with all methods discussed in this paper.

Table XI. Prediction accuracy of 3 support vector machine methods.

Support vector machine methods	Accuracy
SVM	81.9%
SVM Scale	86.1%
SVM Parameter	88.0%

Decision Tree achieves the lowest accuracy with the prediction of only 50.6%, and that is just better than guessing heads or tails when flipping a coin. For all other methods, the accuracy results are above 70%, which is acceptable. Association Rules achieve 76% and 82% accuracy with 1% support and 80% confidence, 1% support and 90% confidence respectively. However, the Association Rules algorithm is the only one without 100% coverage, which makes this algorithm less ideal for the task in this paper. Among our implementation algorithms, Naïve Bayes Laplace archives the highest accuracy of 85.7%. Compared to the Support Vector Machine, the accuracy of Naïve Bayes Laplace beat the original SVM method (85.5% compared to 81.9%).

However, the other two SVM methods generated even better results, especially SVM Parameter with an accuracy of 88%. Based on Figure 6, Naïve Bayes Laplace is able to predict more wines achieving 90+ scores. With an accuracy of 85.7%, it is a successful achievement. In summary, the results suggest that Naïve Bayes Laplace might be the most suitable White-box classification algorithm for understanding wine judges and evaluating the consistency.

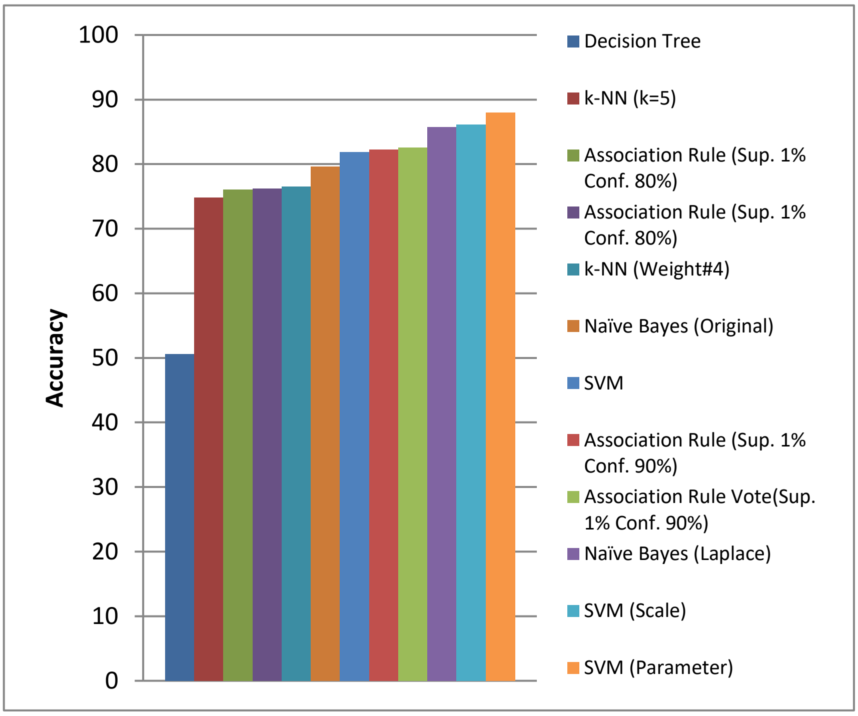

Fig. 6. The comparison chart between Decision Tree, Association Rule, Naïve Bayes (3 methods), k-NN (3 methods), and Support Vector Machine (3 methods) classifications.

Table XII. Sensitivity and Specificity on selected algorithms.

	Sensitivity	**Specificity**
k-NN (k=5)	50.2%	92.4%
Naïve Bayes Laplace	78.6%	92.8%
SVM	73.4%	93.4%

Besides the accuracy comparison, we also worked on the sensitivity (true positive / all positives) and specificity (True Negative / all negatives) comparison on k-NN, Naïve Bayes and SVM algorithms in Table XII. Among all selected algorithms, Naïve Bayes with Laplace gives the best sensitivity values; SVM generates the best specificity values. It is interesting to see that specificity is always higher than 90% in all selected algorithms, while sensitivity is usually lower than 80%. This scenario shows that the prediction of the wine lower than 90 points is easier than the prediction of the wine higher than 90 points.

5 Conclusion and Future Works

White-box classification algorithms are proposed to understand wine judges and evaluate their consistency in this paper. Multiple algorithms with improvements are tested and evaluated in a new data science research domain: Wineinformatics. The overall comparison in Figure 6 shows that Naïve Bayes Laplace is possibly the most suitable white box classification algorithm for understanding wine judges and evaluating their consistency. SVM, a typical Black-box algorithm is also included for comparison purposes.

According to Table XII, improving sensitivity is probably one of the most important and straight forward research problems. Being able to correctly predict when a wine receives a 90+ point will be very useful. In this paper, we evaluate Wine Spectator as a composite professional wine review source. Since Wine Spectator has ten reviewers, each reviewer puts his or her initials at the end of each review. Another important future work is to evaluate every reviewer and possibly rank them according to their consistency. Dimension selection is probably the next most important future work. The dataset in Wineinformatics is considered a high dimensional sparse binary dataset. It is obvious that not all attributes contribute equal weight to the classification process. Identifying significant attributes may further improve the prediction accuracy. Last but not least, multi-source and multi-label techniques can be applied into Wineinformatics. Our testing included using one source, Wine Spectator Magazine wine reviews. There is more than one reliable source for wine reviews. A single wine can be reviewed by more than one source. We suggest using a dataset that includes multiple sources for each wine for its review data.

References

[1] Kantardzic, Mehmed. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, 2011.

[2] International Organization of Wine and Vine [online]. http://www.oiv.int/oiv/cms/index?lang=en.

[3] Global econoic viticiniculture data. Available: http://www.oiv.int/public/medias/4888/en-communiqu-depresse-octobre-2016.pdf

[4] Sun, Li-Xian, Klaus Danzer, and Gabriela Thiel. "Classification of wine samples by means of artificial neural networks and discrimination analytical methods." Fresenius' journal of analytical chemistry 359.2 (1997): 143-149.

[5] Yang, Nan. Quality differentiation in wine markets. Washington State University, 2010

[6] Masset, P., Weisskopf, J. P., & Cossutta, M. (2015). Wine tasters, ratings, and en primeur prices. Journal of Wine Economics, 10(01), 75-107.

[7] Karl Storchmann, "Introduction to the Issue", Journal of Wine Economics, 10(01), 1-3

[8] Bodington, J. C. (2015). Evaluating wine-tasting results and randomness with a mixture of rank preference models. Journal of Wine Economics, 10(01), 31-46.

[9] Stuen, E. T., Miller, J. R., & Stone, R. W. (2015). An analysis of wine critic consensus: A study of Washington and California wines. Journal of Wine Economics, 10(01), 47-61.

[10] Lewis, David D. "Naïve (Bayes) at forty: The independence assumption in information retrieval." Machine learning: ECML-98. Springer Berlin Heidelberg, 1998. 4-15.

[11] Magerman, David M. "Statistical decision-tree models for parsing." In Proceedings of the 33rd annual meeting on Association for Computational Linguistics, pp. 276-283. Association for Computational Linguistics, 1995..

[12] Arias - Bolzmann, Leopoldo, Orkun Sak, Andres Musalem, and Len Lodish. "International Journal of Wine Marketing." Emerald Insight. International Journal of Wine Marketing, 1989. Web. 07 Apr. 2014

[13] Villiers, André De, Phillipus Alberts, Andreas G.j. Tredoux, and Hélène H. Nieuwoudt. "Analytical Techniques for Wine Analysis: An African Perspective; a Review." Analytica Chimica Acta 730 (2012): 2-23.

[14] B. Chen. Wine Attributes [online]. Available http://www.cs.gsu.edu/~cscbecx/Wine%20Informatics.htm. File: Wine_Wheel_01242014.dat

[15] Wine Spectator [online]. http://www.winespectator.com

[16] Wine Spectator (2011) Top 100 [online]. Available: http://www.winespectator.com/display/show?id=45906.

[17] eRobertParker.com. A glossary of Wine Terms [online]. Available: http://www.erobertparker.com/info/glossary.asp.

[18] Sutton, Oliver. "Introduction to k Nearest Neighbour Classification and Condensed Nearest Neighbour Data Reduction." (2012).

[19] Altman, N. S. "An introduction to kernel and nearest-neighbor nonparametric regression." The American Statistician (1992): 46.

[20] Kamber Micheline and Han Jiawei. Data Mining: Concepts and Techniques, 2nd ed. Morgan Kaufmann, 2005.

[21] Hsu, Chih-Wei, Chih-Chung Chang and Chih-Jen Lin. A Practical Guide to Support Vector Classification. Taipei, 19 March 2015.

[22] Friedl, Mark A., and Carla E. Brodley. "Decision tree classification of land cover from remotely sensed data." Remote sensing of environment 61, no. 3 (1997): 399-409.

[23] Chen, Bernard, Rhodes, C., Crawford, A., Hambuchen, L. Wine Informatics: Applying Data Mining on Wine Sensory, Accepted by 2014 Workshop on Domain Driven Data Mining (DDDM 2014).
[24] Agrawal, Rakesh, Tomasz Imieliński, and Arun Swami. "Mining association rules between sets of items in large databases." In ACM SIGMOD Record, vol. 22, no. 2, pp. 207-216. ACM, 1993.
[25] Chemical analysis of grapes and wine: techniques and concepts. Campbelltown, Australia: Patrick Iland wine promotions, 2004.
[26] Cortez Paulo, António Cerdeira, Fernando Almeida, Telmo Matos, and José Reis. Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems 47, no. 4 (2009): 547-553.
[27] Ishibuchi Hisao, Tomoharu Nakashima, and Manabu Nii. Classification and modeling with linguistic information granules: advanced approaches to linguistic data mining. Springer, 2005.
[28] Ishibuchi, Hisao, and Takashi Yamamoto. "Rule weight specification in fuzzy rule-based classification systems." Fuzzy Systems, IEEE Transactions on 13, no. 4 (2005): 428-435.
[29] Olkin, I., Lou, Y., Stokes, L., & Cao, J. (2015). Analyses of wine-tasting data: a tutorial. Journal of Wine Economics, 10(01), 4-30.