The Thirtieth AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI-18)

Investigating the Role of Ensemble Learning in High-Value Wine Identification

Luigi Portinale
Computer Science Institute, DiSIT
Università del Piemonte Orientale
Alessandria, Italy

Monica Locatelli
Department of Pharmaceutical Sciences,
Università del Piemonte Orientale
Novara, Italy

Abstract

We tackle the problem of authenticating high value Italian wines through machine learning classification. The problem is a seriuos one, since protection of high quality wines from forgeries is worth several million of Euros each year. In a previous work we have identified some base models (in particular classifiers based on Bayesian network (BNC), multi-layer perceptron (MLP) and sequential minimal optimization (SMO)) that well behave using unexpensive chemical analyses of the interested wines. In the present paper, we investigate the role of esemble learning in the construction of more robust classifiers; results suggest that, while bagging and boosting may significantly improve both BNC and MLP, the SMO model is already very robust and efficient as a base learner. We report on results concerning both cross validation on two different datasets, as well as experiments with models trained with the above datasets and tested with a dataset of potentially fake wines; this has been synthesized from a generative probabilistic model learned from real samples and expert knowledge. Results open new opportunities in the wine fraud detection activity, which is of primary importance in the figth against the destabilization of the wine market worldwide.

Introduction

The quality and safety profiles of fine wines represent a specific case of the notion of *food integrity*, because of the very high value of a single bottle, and because of the complex chemical profile, requiring specific and robust methods for their univocal authentication. Although specific regulations exist in this matter, quality wines are highly subjected to adulteration. This triggers destabilization of the wine market, with an estimated impact of about 7% of the whole market value, and causing an economical impact estimated to be several million of Euros (Holmberg 2010; Wajsman, Burgos, and Davies 2016). The detection of adulterations is an official task of wine quality control and consumer protection. Specialized analytical methods, based on nuclear magnetic resonance or isotope ratio mass spectrometry exist. However, they are both time consuming and very expensive to undertake; moreover, these methods require high level of specialization and very large data sets. Non analytical approaches like olograms, trasponder systems or QR codes only partially address the problem of wine authenticity. For these reasons, several attempts have been performed, in order to exploit standard analytical chemistry procedures to describe a wine proof-of-identity (Arvanitoyannis et al. 1999; Versari et al. 2014; Marini et al. 2006). Recently, also machine learning techniques have gained attention, due to the fact that classification methodologies can be applied to learn models from samples described through features derived and measured by means of standard chemical analyses (Acevedo et al. 2013; Gòmez-Meire et al. 2014).

In a previous work, we discussed how well-established state-of-the-art supervised machine learning methods can be suitably adopted to fullfill the task of controlling specific wine adulterations (Arlorio et al. 2015). In particular, we referred to the framework of the TRAQUASWINE (TRAceability, QUAlity and Safety of wine) project[1], having as a major goal the authenticity assessment and the protection against fake versions of some of the highest quality (and often top priced) *nebbiolo-based* Italian wines like *Barolo, Barbaresco* and *Gattinara*. We evaluated the application of several classifiers to this task, by considering different datasets. The interpretation of the results suggests an effective role of classification techniques, based on standard chemical profiling, with the emergence of three main models: a Bayesian Network Classifier (BNC), a Multi-layer Perceptron (MLP) and a kernel-based classifier, based on the Sequential Minimal Optimization algorithm (SMO). Given the results obtained using the above "base" learners, a natural question arising is whether the introduction of an *ensemble learning* approach could increase the performance of such base learners, and to provide better results for the target application. In the present paper, we will discuss the results of such an introduction. All the reported experiments have been performed by using the WEKA tool (Hall et al. 2009).

Datasets Production

We collected wine samples (commercial wines) from nine different wineries in Piedmont, Italy with certified wine's

[1]Funded by the European Union, Italian Economy and Finance Ministry, and Regione Piemonte (POR685 FESR 2007/2013 grants).

origin and identity. Eight different types of wines at different aging degrees were considered: *Barolo* (BRL), *Barbaresco* (BRB), *Gattinara* (GAT), *Langhe* (LAN), *Sizzano* (SIZ), *Roero* (ROE), *Nebbiolo d'Alba* (NEB) and *Ghemme* (GHE). Moreover, a set of experimental wines (model wines) has been specifically prepared. Among them, 12 samples were prepared without nebbiolo grape; we labeled them as NON (NO Nebbiolo) and they were used to model possible fakes and forgeries, in particular with respect to BRL and BRB. In addition 10 samples of model wines were produced to model incorrect (i.e., not allowed by the disciplinary of production) blends for GAT wine, and we labeled them as BLE (Blend). The class distribution is reported in Figure 1. The

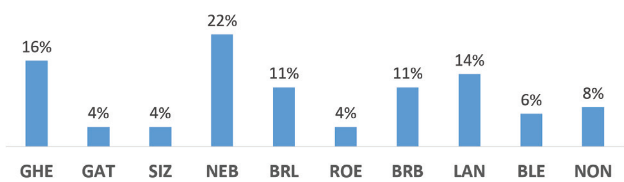

Figure 1: Class (Wine) distribution in the real dataset

study has been organized as follows: we selected some of the most valued nebbiolo-based wines as the high-quality classes to be protected from fakes, and in particular BRL, BRB and GAT; we also selected LAN and BLE as control wines, the former to simulate not allowed blend for BRL and BRB, and the latter for GAT. We finally reserved the 12 samples of NON wines as additional control wines, to test the response of the learned models with respect to simulated fake wines with the absence of nebbiolo grape. We obtained a total of 158 samples with 10 possible classes. Concerning the dataset attributes, wine samples were characterized for their phenolic composition, through a spectrophotometric and chromatographic characterization (Portinale et al. 2017); the performed measurements finally resulted in a set of 13 continuous features reported in Table 1. Missing values were present in about 20% of *cResv* measurements (and in a few other spots).

Classification Results

We considered two different datasets: T1 composed by 158 instances of the 10 classes of wines previously described, and T2 which is obtained from T1 by removing the instances of class NON (resulting in 146 instances). The aim was to verify differences in the performance of the classifiers when possible fake wines without nebbiolo grape are considered in the training or not[2].

Base Learner Classification

We first evaluated the classification performance of the three models previously mentioned: BNC, a Bayesian network classifier learned through the K2 algorithm (Cooper and Herskovits 1992); MLP, a multi-layer perceptron with one hidden layer of $n = \frac{f+c}{2}$ hidden units (being f the number of features and c the number of classes); SMO, a kernel-based classifier based on Sequential Minimal Optimization (Platt 1999) with a Pearson Universal Kernel (PUK) having Lorentzian peak shape (Üstün, Melssen, and Buydens 2006) and with Platt scaling, in order to get a probability distribution over the classes (Platt 2000)[3]. These models have been chosen as the most promising ones after a preliminary experimental phase involving other potential classifiers (Arlorio et al. 2015).

Results for base learner classification are reported in Table 2 and Table 3 for dataset T1 and T2 respectively. They show results about 10-fold cross validation on the datasets, by reporting general accuracy (**Acc**), KAPPA statistics (k) and weighted (with respect to the number of instances) average area under ROC (**AUC**); moreover for each high-quality and control class, the number of predictions is also shown (e.g., in Table 2/BNC, LAN was predicted 16 times as LAN, and 6 times as NEB). We can notice that general classification performance of the tested classifiers is rather good, with some problems evidenced with BNC. However, we are particularly interested in evaluating the classification behavior with respect to high quality and control wines. To this end, no control wine is predicted as high quality and vice versa, in all the tested classifiers. In particular, for dataset T1, there is always a perfect recognition on NON wines (absent from T2). Moreover, given the grape/terroir features of the wines, mispredictions can be quite often justified by similar wines in grape composition and from close geographical areas (e.g., BRL, BRB and NEB from south Piedmont or BLE and GHE from north Piedmont). Remarkable is the performance of SMO on both T1 and T2, with a perfect classification of GAT and LAN, and almost perfect for the other classes (perfect on BRL on T2).

Ensemble Classification: bagging and boosting

Given the results obtained from base learner classification, we have then investigated the potential benefit of an ensemble learning approach (especially for BNC and MLP). Two general kinds of ensembles were taken into account: the *bagging/boosting* of each base classifier, and the *stacking/voting* of the whole set of base classifiers. This section reports the results for the first kind of ensemble, while next section shows the results concerning stacking/voting.

Bagging implementation was set with 10 iterations using all the instances in the dataset; the chosen boosting approach has been AdaBoost with 10 iterations and explicit instance reweighting. Results related to 10-fold cross validation on datasets T1 and T2 are shown from Table 4 to Table 9.
Concerning BNC, we notice an improvement in the general accuracy both on T1 and T2; with respect to predictions, AdaBoost provides a significant improvement on T2. Similar considerations hold for MLP, where AdaBoost provides better and significant improvements on predictions, especially on T1. A peculiar case is SMO, where both bagging and boosting do not result in any significant deviation from the

[2]We remark that it is can be hard in general, to have a precise idea of the chemistry of fakes in advance.

[3]Hyper-parameters for such models have been chosen through cross-validation.

Feature	Acronym	Feature	Acronym
Total Polyphenols	TP	Total Tannins	TT
Peonidin-3-O-glucoside	Pn-3-glc	Malvidin-3-O-glucoside	Mv-3-glc
Delphinidin-3-O-glucoside (perc)	PDn-3-glc	Peonidin-3-O-glucoside (perc)	PPn-3-glc
Malvidin-3-O-glucoside (perc)	PMv-3-glc	Caffeic Acid	Caff
Ferulic Acid	Fer	Kaempferol-3-O-glucoside	Kae-3-glu
Myricetin	Mir	Coutaric Acid (perc)	PCout
cis-Resveratrol	cResv		

Table 1: Features names and acronyms

(a) BNC	**Acc: 73% k: 0.69 AUC: 0.96**
GAT	GAT (**4**); GHE (**1**); NEB (**1**)
BRL	BRL (**13**); GHE (**2**); BRB (**3**)
BRB	BRB (**13**); NEB (**2**); BRL (**2**); ROE (**1**)
LAN	LAN (**16**); NEB (**6**)
BLE	BLE (**7**); GHE (**1**); NEB (**1**); ROE (**1**)
NON	NON (**12**)

(b) MLP	**Acc: 89% k: 0.86 AUC: 0.97**
GAT	GAT (**6**)
BRL	BRL (**14**); GHE (**1**); BRB (**2**); ROE (**1**)
BRB	BRB (**14**); NEB (**2**); BRL (**2**)
LAN	LAN (**22**)
BLE	BLE (**8**); NEB (**1**); LAN (**1**)
NON	NON (**12**)

(c) SMO	**Acc: 93% k: 0.92 AUC: 0.997**
GAT	GAT (**6**)
BRL	BRL (**18**)
BRB	BRB (**16**); BRL (**2**)
LAN	LAN (**22**)
BLE	BLE (**7**); GHE (**3**)
NON	NON (**12**)

Table 2: Base learner classification: dataset T1

base learner case (with AdaBoost providing exactly the same cross validation results)

Ensemble Classification: stacking and voting

We also evaluated the ensemble of different types of base classifiers, by investigating approaches based on both stacking and voting. In the case of stacking, the main design decision regards the choice of the meta-learner, while in the case of voting, the main choice concerns the aggregation function. In our case, the most relevant results have been obtained by using a 1-NN (*nearest neighbor*) and a *Logistic Regression* meta-learner in the case of stacking, and the *average* aggregation function in the case of voting. Table 10 and Table 11 show the results obtained with a 10-fold cross validation on datasets T1 and T2 respectively (Stack1NN refers to the 1-NN meta-learner and StackLog to the logistic regression one). We notice that StackLog is the ensemble with the worst performance both in terms of general accuracy and predictions, while Stack1NN and Voting provides comparable results that are very close to those of SMO on both datasets.

(a) BNC	**Acc: 79% k: 0.75 AUC: 0.97**
GAT	GAT (**5**); BRL (**1**)
BRL	BRL (**13**); GAT/NEB/ROE (**1**); BRB (**2**)
BRB	BRB (**14**); NEB (**1**); BRL (**2**); ROE (**1**)
LAN	LAN (**20**); NEB (**2**)
BLE	BLE (**8**); GHE (**1**); NEB (**1**)

(b) MLP	**Acc: 81% k: 0.79 AUC: 0.92**
GAT	GAT (**6**)
BRL	BRL (**14**); GHE (**1**); BRB (**1**); ROE (**2**)
BRB	BRB (**12**); NEB (**2**); BRL (**3**); ROE (**1**)
LAN	LAN (**22**)
BLE	BLE (**9**); GHE (**1**)

(c) SMO	**Acc: 89% k: 0.87 AUC: 0.98**
GAT	GAT (**6**)
BRL	BRL (**16**); BRB (**2**)
BRB	BRB (**16**); BRL (**2**)
LAN	LAN (**22**)
BLE	BLE (**8**); GHE (**2**)

Table 3: Base learner classification: dataset T2

Testing Fakes

To complete our evaluation, we verified the different behavior of the considered classifiers when presented with a potential fake wine without nebbiolo grape (the most probable fake). In order to perform this test, we started with the 12 samples of NON model wines previously described, and performed an artificial generation of a synthetic dataset of 1200 instances of potential fakes. Data generation has been implemented by first learning a *Linear Gaussian Bayesian Network* (LGBN) from the available samples and some expert background knowledge (figure 2)[4]. This allowed us to have a generative model for continuous features, from which to produce the synthetic dataset.

We perfomed two different types of test. In the first one, we considered T1 as the tranining set and we used the synthetic dataset as the test set. This allowed to test the different classifiers once trained with some potentially fakes without nebbiolo grape. Results are summarized in Table 12. In the second test, we considered the dataset T2 as training set for learning, and the synthetic dataset, augmented with

[4]The networks was learned using the PC algorithm (Spirtes, Glymour, and Scheines 1993) implemented in SMILE (www.bayesfusion.com).

Bagging	**Acc: 85% k: 0.82 AUC: 0.99**
GAT	GAT (**4**); GHE (**1**); BRL (**1**)
BRL	BRL (**14**); BRB (**2**); GHE (**2**)
BRB	BRB (**17**); BRL (**1**)
LAN	LAN (**19**); NEB (**3**)
BLE	BLE (**7**); GHE (**2**); NEB (**1**)
NON	NON (**11**); LAN (**1**)

AdaBoost	**Acc: 86% k: 0.84 AUC: 0.98**
GAT	GAT (**5**); NEB (**1**)
BRL	BRL (**13**); GHE (**1**); BRB (**2**); NEB (**2**)
BRB	BRB (**15**); NEB (**1**); BRL (**1**); ROE (**1**)
LAN	LAN (**20**); NEB (**2**)
BLE	BLE (**8**); GHE (**1**); LAN (**1**)
NON	NON (**12**)

Table 4: BNC: ensemble classification on dataset T1

Bagging	**Acc: 84% k: 0.81 AUC: 0.91**
GAT	GAT (**5**); BRL (**1**)
BRL	BRL (**15**); BRB (**1**); NEB (**1**); GAT (**1**)
BRB	BRB (**16**); BRL (**1**); ROE (**1**)
LAN	LAN (**19**); NEB (**3**)
BLE	BLE (**8**); GHE (**1**); NEB (**1**)

AdaBoost	**Acc: 87% k: 0.85 AUC: 0.98**
GAT	GAT (**6**)
BRL	BRL (**17**); NEB (**1**)
BRB	BRB (**15**); NEB (**2**); ROE (**1**)
LAN	LAN (**20**); NEB (**2**)
BLE	BLE (**9**); GHE (**1**)

Table 5: BNC: ensemble classification on dataset T2

the original 12 NON samples, as the test set. In this case, the classifiers have been learned without training on potential fakes without nebbiolo grape. Corresponding results are shown in Table 13. In these tables we report the predictions of each classifier, accounting for more than 95% of the total. We also consider the so called *fake probability*, that is the probability of predicting a high-quality wine from a fake. They are reported (as percentage values) by considering a 98% confidence interval of the average probability of predicting either BRL or BRB or GAT wines (symbols / means a negligible value, i.e., if the average probability is less than 1%). Unsurprisingly, test with training set T1 provides better results than test with training set T2; experiments confirm several findings from cross validation, and in particular, the very good performances of the base learner SMO and of ensembles based on Stack1NN and Voting, as well as the big role of bagging for BNC and of boosting for both BNC and MLP. With training set T1, SMO based methods provide almost perfect prediction; this occurs for Stack1NN and Voting as well. However, Voting has a non negligible average probability of predicting a high quality wine (BRB), even if this value is only between 1.8% and 2.4% at the 98% confidence level. In addition to the results shown in Table 12, experiments showed some additional merits of kernel-based methods, since they predict NON with an average 98% probability in case of SMO and AdaBoostSMO, and with 81%

Bagging	**Acc: 89% k: 0.87 AUC: 0.99**
GAT	GAT (**6**)
BRL	BRL (**13**); BRB (**2**); NEB (**2**); GAT (**6**)
BRB	BRB (**16**); BRL (**2**)
LAN	LAN (**22**)
BLE	BLE (**8**); GHE (**1**); LAN (**1**)
NON	NON (**12**)

AdaBoost	**Acc: 94% k: 0.93 AUC: 0.98**
GAT	GAT (**6**)
BRL	BRL (**17**); GHE (**1**)
BRB	BRB (**16**); BRL (**2**)
LAN	LAN (**22**)
BLE	BLE (**10**)
NON	NON (**12**)

Table 6: MLP: ensemble classification on dataset T1

Bagging	**Acc: 82% k: 0.79 AUC: 0.96**
GAT	GAT (**6**)
BRL	BRL (**14**); GHE (**2**); SIZ (**1**); NEB (**1**)
BRB	BRB (**13**); NEB (**2**); BRL (**3**)
LAN	LAN (**22**)
BLE	BLE (**8**); GHE (**1**); LAN (**1**)

AdaBoost	**Acc: 90% k: 0.89 AUC: 0.97**
GAT	GAT (**6**)
BRL	BRL (**16**); SIZ (**2**)
BRB	BRB (**15**); NEB (**1**); BRL (**2**)
LAN	LAN (**22**)
BLE	BLE (**10**)

Table 7: MLP: ensemble classification on dataset T2

probability in case of BagSMO. In general, NON has always a non-negligible probability of prediction even when it is not the predicted class (and often is the second predicted class). From the negative side, BagBN has some predictions of BRB (9) with an average probability of prediction of about 35%.

By considering training set T2, NON wines cannot be recognized, since there is no explicit class in the learned models for them. The most similar class that can be used as a target is LAN, with BLE as a second choice. Of course, we would like to avoid predictions of high quality wines, but also of NEB which is a wine with 100% of nebbiolo grape. By looking at the results in Table 13, very often the tested wine is

Bagging	**Acc: 92% k: 0.91 AUC: 0.99**
GAT	GAT (**6**)
BRL	BRL (**18**)
BRB	BRB (**17**); BRL (**1**)
LAN	LAN (**22**)
BLE	BLE (**8**); GHE (**2**)
NON	NON (**12**)

Table 8: SMO: ensemble classification on dataset T1 (AdaBoost on SMO provides the same results as those of Table 2 (c))

Bagging	**Acc: 88%** k**: 0.86 AUC: 0.93**
GAT	GAT (**6**)
BRL	BRL (**17**); NEB (**1**)
BRB	BRB (**16**); BRL (**2**)
LAN	LAN (**22**)
BLE	BLE (**7**); GHE (**2**); LAN (**1**)

Table 9: SMO: ensemble classification on dataset T2 (AdaBoost on SMO provides the same results as those of Table 3 (c))

Stack1NN	**Acc: 90%** k**: 0.88 AUC: 0.97**
GAT	GAT (**6**)
BRL	BRL (**18**)
BRB	BRB (**15**); NEB (**1**); BRL (**2**)
LAN	LAN (**21**); NEB (**1**)
BLE	BLE (**5**); GHE (**4**); SIZ (**1**)
NON	NON (**12**)
StackLog	**Acc: 82%** k**: 0.80 AUC: 0.80**
GAT	GAT (**6**)
BRL	BRL (**16**); NEB (**1**); BRB (**1**)
BRB	BRB (**13**); NEB (**4**); BRL (**1**)
LAN	LAN (**20**); NEB (**2**)
BLE	BLE (**9**); NEB (**1**)
NON	NON (**11**); NEB (**1**)
Voting	**Acc: 93%** k**: 0.92 AUC: 0.99**
GAT	GAT (**6**)
BRL	BRL (**17**); BRB (**1**)
BRB	BRB (**16**); BRL (**2**)
LAN	LAN (**22**)
BLE	BLE (**8**); GHE (**1**); NEB (**1**)
NON	NON (**12**)

Table 10: Stacking/Voting classification: dataset T1

confused with a NEB, even if in such situations we have noticed that LAN is often the second prediction in order of probability (and with a very close prediction probability wrt NEB). This is the case for SMO based methods, with bagging increasing the predictions of LAN, but by introducing more probable predictions of BRL and BRB as well.

Conclusions

We have reported the results of a study exploiting classification for the assessment of the authenticity of some high-value Italian wines. Forgeries about this kind of wines are economically very relevant, but standard fake discovering techniques usually employ very expensive instrumental analyses; we have proposed to exploit simple chemical analyses, coupled with off-the-shelf machine learning methodologies. We have investigated the role of ensemble learning in the considered application context. Experiments have recognized an active role of classifiers like BNC, MLP and SMO. The introduction of ensemble techniques results to be beneficial for BNC and MLP, with boosting being in general better than bagging. However, kernel-based classification based on SMO seems to be quite insensitive to either bagging or boosting in this application, and even combination of the different base learners through stacking or voting does not seem to significantly improve SMO. The positive evaluation of the described classification methods for an application requiring only standard lab analyses, definitely opens new opportunities in the wine fraud detection activity, which is considered of primary importance in the figth against the destabilization of the wine market worldwide.

Stack1NN	**Acc: 88%** k**: 0.86 AUC: 0.97**
GAT	GAT (**6**)
BRL	BRL (**14**);BRB (**4**)
BRB	BRB (**16**); BRL (**2**)
LAN	LAN (**21**); NEB (**1**)
BLE	BLE (**10**)
StackLog	**Acc: 81%** k**: 0.78 AUC: 0.90**
GAT	GAT (**5**); GHE (**1**)
BRL	BRL (**14**); NEB (**1**); BRB (**2**); GHE (**1**)
BRB	BRB (**15**); NEB (**1**); BRL (**2**)
LAN	LAN (**21**); NEB (**1**)
BLE	BLE (**7**); NEB (**1**); BRL (**1**); ROE (**1**)
Voting	**Acc: 90%** k**: 0.88 AUC: 0.99**
GAT	GAT (**6**)
BRL	BRL (**18**)
BRB	BRB (**15**); BRL (**2**); ROE (**1**)
LAN	LAN (**22**)
BLE	BLE (**9**); GHE (**1**)

Table 11: Stacking/Voting classification: dataset T2

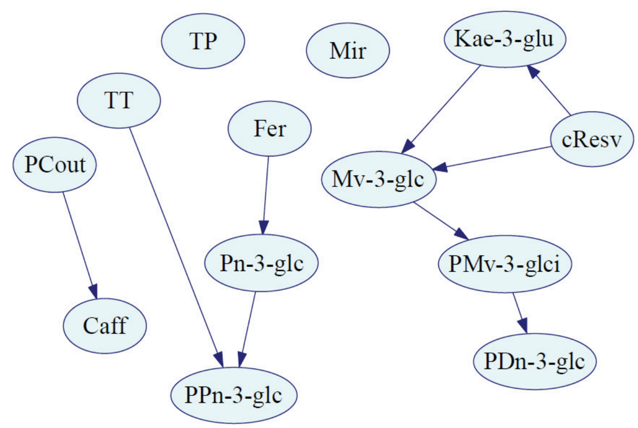

Figure 2: LGBN structure for generating the test set

References

Acevedo, F.; Nez, J.; Maldonado, S.; Domínguez, E.; and Narváez, A. 2013. Classification of wines produced in specific regions by UV-visible spectroscopy combined with support vector machines. *Journal of Agricultural and Food Chemistry* 55:6842–6849.

Arlorio, M.; Coisson, J.; Leonardi, G.; Locatelli, M.; and Portinale, L. 2015. Exploiting data mining for authenticity assessment and protection of high-quality Italian wines from

Classifier	Predictions	Fake prob. (perc.)
BNC	NON (**85%**); LAN (**9%**); ROE (**3%**)	/
MLP	NON (**74%**); LAN (**17%**); LAN (**4%**)	BRB:**[4.7,6.2]**
SMO	NON (**99.3%**)	/
AdaBoost:BNC	NON (**95%**); LAN (**4%**)	/
AdaBoost:MLP	NON (**87%**); LAN (**12%**)	/
AdaBoost:SMO	NON (**99.3%**)	/
Bagging:BNC	NON (**88%**); LAN (**9%**)	BRB:**[2.7,3.4]**
Bagging:MLP	NON (**86%**); LAN (**14%**)	BRB:**[2.4,2.9]**; BRL:**[1, 1.3]**
Bagging:SMO	NON (**93%**); LAN (**6%**)	/
Stack1NN	NON (**99%**)	/
StackLog	NON (**77%**); LAN (**10%**); BRL (**7%**); NEB (**5%**);	BRL: **[5.4,8.6]**
Voting	NON (**97%**)	BRB: **[1.8,2.4]**

Table 12: Testing NON: training set T1

Classifier	Predictions	Fake prob. (perc.)
BNC	LAN (**29%**); BLE (**27%**); NEB (**19%**); ROE (**19%**); BRB (**4%**)	GAT:**[1.6,2.4]**; BRL:**[1.7,2.5]**; BRB:**[6.3,8]**
MLP	NEB (**42%**); LAN (**34%**); BLE (**9%**); BRL (**8%**); ROE (**6%**);	BRL:**[7.1,9.9]**;BRB:**[1.8,2.9]**
SMO	LAN (**69%**); NEB (**31%**)	/
AdaBoost:BNC	LAN (**66%**); NEB (**22%**); BLE (**10%**)	/
AdaBoost:MLP	LAN (**48%**); NEB (**44%**); BRB (**6%**)	BRB:**[5,7.9]**
AdaBoost:SMO	LAN (**69%**); NEB (**31%**)	/
Bagging:BNC	LAN (**84%**); NEB (**9%**); BLE (**4%**)	BRB:**[3.6,4.4]**
Bagging:MLP	LAN (**48%**); NEB (**44%**); ROE (**5%**)	BRB:**[2.4,2.9]**; BRL:**[4.2, 4.9]**; BRB:**[7, 7.9]**
Bagging:SMO	LAN (**77%**); NEB (**23%**)	BRL: **[1.9,2.3]**; BRB: **[2.1,2.5]**
Stack1NN	NEB (**52%**); LAN (**46%**)	/
StackLog	NEB (**40%**); BRB (**31%**); LAN (**23%**); BRL (**4%**);	BRL: **[2.6,5.1]**; BRB: **[28.4,34.6]**
Voting	LAN (**54%**); NEB (**42%**)	BRL: **[3.4,4.4]**; BRB: **[3.3,3.9]**

Table 13: Testing NON: training set T2

Piedmont. In *Proc. 21th Int. Conf. on Knowledge Discovery and Data Mining (KDD'15)*. 1671–1680.

Arvanitoyannis, I.; Katsota, M.; Psarra, E.; Soufleros, E.; and Kallithraka, S. 1999. Application of quality control methods for assessing wine authenticity: Use of multivariate analysis (chemometrics). *Trends in Food Science and Technology* 10:321–336.

Cooper, G., and Herskovits, E. 1992. A bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9(4):309–347.

Gòmez-Meire, S.; Campos, C.; Falqué, E.; Dìaz, F.; and Fdez-Riverola, F. 2014. Assuring the authenticity of northwest Spain white wine varieties using machine learning techniques. *Food Research International* 60:230–240.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; and Witten, I. 2009. The WEKA data mining software: An update. *SIGKDD Explorations* 11(1):10–18.

Holmberg, L. 2010. Wine fraud. *International Journal of Wine Research* 2010(2):105–113.

Marini, F.; Bucci, R.; Magrì, A. L.; and Magrì, A. D. 2006. Authentication of italian CDO wines by class-modeling techniques. *Chemometrics and Intelligent Laboratory Systems* 84(1):164–171.

Platt, J. 1999. Fast training of support vector machines using sequential minimal optimization. In Schölkopf, B.; Burges, C.; and Smola, A., eds., *Advances in Kernel Methods*. MIT Press. 185–208.

Platt, J. 2000. Probability for SV machines. In Smola, A.; Batlett, P.; Schölkopf, B.; and Schuurmans, D., eds., *Advances in Large Margin Classifiers*. MIT Press. 61–74.

Portinale, L.; Arlorio, M.; Coisson, J.; Leonardi, G.; Travaglia, F.; and Locatelli, M. 2017. Authenticity assessment and protection of high-quality nebbiolo-based Italian wines through machine learning. *Chemometrics and Intelligent Laboratory Systems* 171:182–197.

Spirtes, P.; Glymour, C.; and Scheines, R. 1993. *Causation, Prediction and Search*. Berlin: Springer Verlag.

Üstün, B.; Melssen, W.; and Buydens, L. 2006. Facilitating the application of support vector regression by using a universal Pearson VII function based kernel. *Chemometrics and Intelligent Laboratory Systems* 81:29–40.

Versari, A.; Laurie, V.; Ricci, A.; Laghi, L.; and Parpinello, G. 2014. Progress in authentication, typification and traceability of grapes and wines by chemometric approaches. *Food Research International* 60:2–18.

Wajsman, N.; Burgos, C. A.; and Davies, C. 2016. The economic cost of IPR infringement in spirits and wine. Technical Report 07/2016, EUIPO Observatory.