

Paladar Artificial: Análise Sensorial para Classificação de Vinhos Aplicando Algoritmo Firework

Gabriel Ferreira Neves (UFPR – Campus Jandaia do Sul) gabrielneves@ufpr.br

Rodrigo Clemente Thom de Souza (UFPR – Campus Jandaia do Sul) thom@ufpr.br

Resumo:

Vinho, bebida milenar e histórica, produzida a partir da fermentação alcoólica das uvas, que fermentam, devido sua constituição química, sem que lhe seja adicionado qualquer reagente. Desde o século 19 existem muitos estudos sobre a bebida mas apenas no século 21 descobriu-se que os vinhos se diferenciam uns dos outros em termos de propriedades químicas e podem ser classificados segundo suas características. Neste contexto, um método bastante utilizado na indústria alimentícia é a análise sensorial, sendo empregada principalmente na produção de bebidas, com objetivo de melhorar gradualmente o produto e verificar a aceitação do consumidor. Existem diversos algoritmos de classificação de dados que com auxílio da análise sensorial buscam reconhecer padrões e prever resultados, dentre eles o *Firework Algorithm*. Este algoritmo explora um determinado espaço de solução selecionando pontos aleatórios, e procurando a solução ótima para uma determinada base de dados tendo como base a explosão de fogos de artifício no céu. O objetivo deste estudo é mostrar que utilizando *Firework Algorithm* de maneira binarizada seria possível criar padrões para classificação de vinhos a partir de suas características descobrindo a qual classe o mesmo pertence ou prevendo qual o melhor vinho quando provado por um especialista. De acordo com os resultados obtidos, o algoritmo mostrara eficiência satisfatória para todas as bases de dados, considerando-se a complexidade do problema e tornando o método proposto bastante promissor.

Palavras chave: Enologia, Análise Sensorial, *Firework Algorithm*.

Artificial Tasting: Computational Sensory Analysis for Wine Classification Applying Firework Algorithm

Abstract

Wine, millennial and historical drink, produced from the alcoholic fermentation of the grapes, which ferment, due to their chemical constitution, without being added any reagent. Since the 19th century there have been many studies on the beverage but only in the 21st century it has been discovered that wines differ from each other in terms of chemical properties and can be classified according to their characteristics. In this context, a method widely used in the food industry is the sensory analysis, being used mainly in the production of beverages, with the objective of gradually improving the product and verifying consumer acceptance. There are several data classification algorithms that, with the aid of sensory analysis, seek to recognize patterns and predict results, among them Firework Algorithm. This algorithm explores a particular solution space by selecting random points, and searching for the optimal solution for a particular database based on the explosion of fireworks in the sky. The objective of this study is to show that using Firework Algorithm in a binarized manner it would be possible to create standards for classifying wines from their characteristics by discovering which class it belongs to or predicting which wine is best when tested by a specialist. According to the obtained results, the algorithm will show satisfactory efficiency for all the databases, considering the complexity of the problem and making the proposed method quite promising.

Key-words: Oenology, Sensory Analysis, Firework Algorithm.

1. Introdução

Vinho, bebida produzida por fermentação alcoólica da uva inteira, esmagada ou em mosto como afirma o artigo 3º da lei 7678/88º (BRASIL, 1988). Presume-se, segundo *Berkowitz* (1996), que o surgimento do vinho tenha ocorrido aproximadamente em 8000 a.C na Turquia, em Damasco na Síria, *Byblos* no Líbano e na Jordânia.

Até meados do século 19, não se sabia muito sobre o processo de fermentação da uva ou de deterioramento do vinho. Tantos os gregos como os romanos bebiam os seus vinhos logo no primeiro ano após o preparo, devido à falta de técnicas para a conservação eficaz. Tais povos costumavam adicionar flavorizantes, como ervas, mel, queijo ou até mesmo sal, para disfarçar o gosto do vinagre que os vinhos possuíam.

O químico Louis Pasteur foi um dos pioneiros no estudo da bebida explicando a origem química da fermentação e os agentes responsáveis pelo processo, segundo o francês “Existe mais filosofia em uma garrafa de vinho que em todos os livros”.

As uvas fermentam, devido a sua constituição química, sem que lhes sejam adicionados açúcares, ácidos, enzimas ou outros nutrientes (Simon et al., 1989). A fermentação das uvas é feita por vários tipos de leveduras que consomem os açúcares presentes nas uvas transformando-os em álcool.

No século 21 sabe-se, com base em estudos, que os vinhos diferenciam-se uns dos outros em termos de propriedades químicas e podem ser classificados segundo suas características, tais como cor, teor de açúcar, variedade da uva, entre outros. Em termos de constituição química, podem ser diferenciados pela quantidade de álcool, ácidos, açúcares, cloretos, óxidos, metais, sulfatos, entre outros.

Caso sejam conhecidos os valores característicos de determinados vinhos, como o acidez ou o teor de açúcar, por exemplo, e a nota que especialistas em vinhos (enólogos e *sommeliers*)

deram para os mesmos, espera-se ser possível avaliar outros vinhos, estimando qual deles seria melhor avaliado por um especialista sem que o mesmo fosse provado. De forma análoga, poderia ser possível identificar cultivares, cepas ou até mesmo a região de procedência de um vinho através das características químicas do mesmo.

Na base de dados *Winequality* foram analisados 1599 vinhos verdes tintos e 4898 vinhos verdes brancos. Na *Wine* foram analisados 178 vinhos italianos, classificados em três classes distintas de acordo com um cultivar da Itália.

Agustini *et al.* (2012) utilizou métodos de análise sensorial para determinar para avaliar as concentrações dos ácidos málico, láctico e tartárico em vinhos comuns elaborados com uva Terci, para verificar a influência destes compostos na acidez e no pH dos vinhos.

Neves *et al.* (2016) utilizou as mesmas bases de dados para mostrar as vantagens de um produtor agrícola de vinhos fazer o uso de Regressão Logística aplicada com *bootstrap* para prever a nota do vinho que ele produz.

O objetivo deste estudo é mostrar que utilizando *Firework Algorithm (FWA)* de maneira binarizada como seletor de atributos podem-se criar padrões para classificação de vinhos a partir de suas características e assim colocá-los em classes e prever qual o melhor vinho quando provado por um especialista.

Assim, através das características de um vinho prever testes de qualidade, confirmação de procedência da uva utilizada, economizar dinheiro com especialistas, pois a possível avaliação pode ser prevista sem o mesmo ser provado.

2. Análise Sensorial

Análise sensorial foi definida como uma disciplina científica usada para evocar, medir, analisar e interpretar reações das características dos alimentos e materiais, como são percebidas pelos sentidos visão, olfato, sabor, tato e audição (AMERINE; PANGBORN; ROESSLER, 1965).

Método comumente utilizado na indústria alimentícia, devido à possibilidade de determinar a aceitabilidade e qualidade dos alimentos com base nos órgãos de sentido humano.

Segundo Araújo (2013), tais as análises podem ser capazes de determinar compostos que influenciam as características sensoriais e nutricionais das uvas.

Em termos sensoriais, pode-se recorrer a análise sensorial, visto que é um método científico aplicado para medir, analisar, interpretar e avaliar características de um produto e o modo como são percebidas pelos sentidos humanos (tato, olfato, paladar, visão). Entre os métodos aplicados para análise sensorial estão os métodos descritivos que caracterizam um produto com relação aos diversos atributos sensoriais percebidos pelo provador, como cor, sabor, odor e aparência.

O principal método descritivo utilizado é a Análise Descritiva Quantitativa (ADQ[®]) Tal metodologia avalia todos os atributos sensoriais presentes no produto, sendo eles: aroma, aparência, sabor e textura, com base em uma escala não estruturada de 9 cm, com termos um pouco aquém dos extremos indicando a intensidade do atributo que está sendo avaliado. Além disso, pode-se aplicar testes afetivos com intuito de avaliar o grau de aceitabilidade de um alimento ou produto pelos consumidores.

Ao aplicar tal método em degustação de vinhos brancos as amostras devem estar entre 12°C e 18°C para vinhos tintos a temperatura ideal das amostras é de 22°C, tais temperaturas garantem que as características físico-químicas dos mesmos não sejam perdidas (Dutcosky, 2011).

3. Firework Algorithm (FWA)

Tendo como base o princípio da explosão de fogos de artifício Tan, Y., & Zhu, Y. (2010) propuseram que o FWA explora um grande espaço de solução selecionando pontos aleatórios, aplicando métricas de distância na esperança de que um ou mais destes fogos produza resultados promissores para na próxima geração realizar uma busca mais concentrada.

A partir da observação das curvas de algumas funções típicas como *Rosenbrock*, *Schwefel*, *Griewank* e *Sphere* os pontos perto do ótimo global e local sempre possuem melhores valores de *fitness*. Assim, controlando a amplitude da explosão e dos fogos diminui-se gradativamente cada *fitness* melhor, levando os fogos a achar o ótimo global e local.

O FWA é composto de 4 partes: operador explosivo, operador de mutação, regra de mapeamento e estratégia de seleção.

3.1. Operador Explosivo

Inicialmente o algoritmo gera N fogos de artifício aleatoriamente, os mesmos produzem faíscas por operações de explosões.

3.1.1. Força de Explosão

Operação central deste operador, simula a maneira real de explosão dos fogos de artifício. Quando um fogo estoura desaparece em segundos gerando várias pequenas explosões ao seu redor. Nesta etapa é necessário calcular o número de faíscas a partir da equação 1 para calcular a amplitude de cada explosão.

$$S_i = m \cdot \frac{Y_{max} - f_{(x_i)} + \varepsilon}{\sum_{i=1}^N (Y_{max} - f_{(x_i)}) + \varepsilon}$$

, onde m é constante, Y_{max} o fitness do pior indivíduo, $f_{(x_i)}$ o fitness de um indivíduo i e ε um valor para que o denominador não seja 0.

A limitação do número de faíscas é dada pela equação 2 por:

$$S_i = \begin{cases} \text{round}(a \cdot m), & \text{se } S_i < a \cdot m \\ \text{round}(b \cdot m), & \text{se } S_i > b \cdot m \\ a < b < 1 \end{cases}$$

, onde as variáveis “a”, “m” e “b” são constantes.

O FWA determina a amplitude e os números dos fogos de artifício de acordo com os melhores valores de aptidão, possibilitando que os fogos produzam mais faíscas dentro de uma amplitude menor para encontrar a solução do problema como visto na figura 1

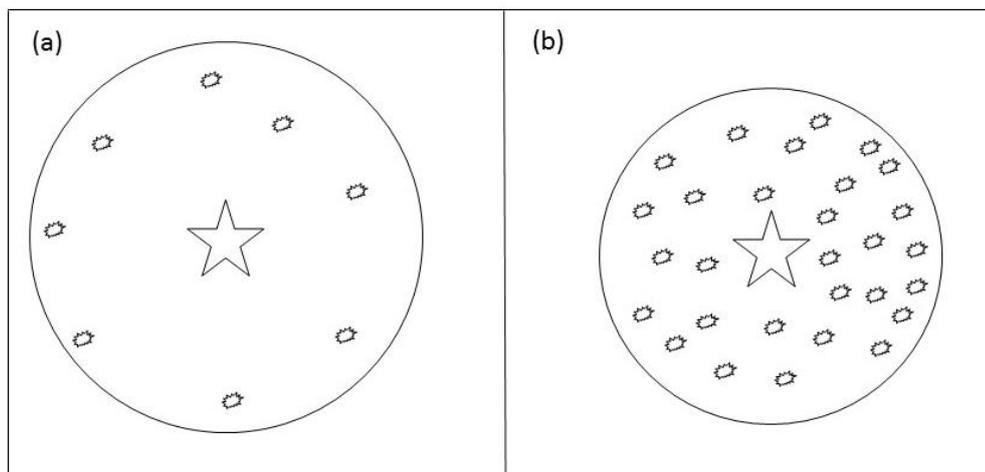


Figura 1 – As boas e más explosões de um fogo de artifício, (a) Explosão ruim (b) Explosão boa. Fonte: Y. Tan, Y. Zhu, *Fireworks algorithm for optimization*, adaptado pelo autor.

3.1.2. Amplitude de Explosão

A amplitude de explosão é dada pela equação 3:

$$A_i = \hat{A} \cdot \frac{f(x_i) - Y_{min} + \varepsilon}{\sum_{i=1}^N (f(x_i) - Y_{min}) + \varepsilon}$$

, onde A_i é a amplitude de cada indivíduo, \hat{A} é constante referente a soma de todas as amplitudes e Y_{min} significa o valor de aptidão do melhor indivíduo.

3.1.3. Operador de Deslocamento

Este operador executa o deslocamento em cada dimensão de um fogo de artifício e é definido pela equação 4:

$$z = \text{round}(d \cdot \text{rand}(0,1))$$

, onde d é a dimensão da localização x e $\text{rand}(0,1)$ corresponde a distribuição dentro de $[0, 1]$.

3.2. Operador de Mutação

Tem objetivo de diferenciar as populações a cada geração de fogos, a mutação gaussiana quando aplicada a partir da escolha de um fogo da população atual gera novas faíscas diferentes das anteriores.

Segundo Tan, Y. (2010) a mutação gaussiana, faz com que as novas faíscas sejam geradas entre os melhores fogos de artifício e aqueles selecionados. Porém, a tal mutação pode produzir faíscas que excedam o espaço viável, nestes casos a regra de mapeamento será aplicada e a faísca será realocada para um novo local no espaço viável.

3.3. Regra de Mapeamento

Garante que todos os indivíduos permaneçam no espaço viável, caso haja algum *outlier* o mesmo será mapeado de volta ao permitido.

3.4. Estratégia de Seleção

Nesta etapa do algoritmo é utilizado o princípio de medida de distribuição euclidiana, dada pela equação 5:

$$R(x_i) = \sum_{j=1}^K d(x_i, j_i) = \|x_i - j_i\|$$

, onde $R(x_i)$ representa a soma das distâncias entre o indivíduo X_i e todos os outros indivíduos. O somatório $j \in K$ significa que a posição j pertence a K , conjunto este que combina faíscas geradas pelo operador de explosão e mutação.

Os indivíduos da próxima geração são escolhidos a partir da equação do modo roleta (6), onde indivíduos com maiores distâncias possuem maior probabilidade de serem selecionados para próxima geração.

$$p(x_i) = \frac{R(x_i)}{\sum_{j \in K} R(x_j)}$$

4. Materiais e Métodos

O presente artigo realiza a classificação dos dados das bases utilizando o algoritmo Naive Bayes embrulhado por um seletor de atributos baseado no algoritmo *FWA*. Classificação de dados é a ação de atribuir classes para as amostras, baseado em suas características podendo assim discriminar em classes dados. Segundo Thom de Souza (2013), não só as quantidades de dados definem a qualidade dos algoritmos, mas também da qualidade dos dados (informações relevantes). A seleção de atributos, por sua vez, trata da escolha dos melhores atributos de predição dentro de um conjunto de dados visando aumentar a acurácia de classificação.

A análise de dados foi realizada Intel core i5 com Windows 10 de 64 bits. “*Winequality*” uma das bases de dados utilizadas, provinda da *University of California Irvine (UCI)* utiliza 11 parâmetros (camadas de entrada) sendo eles: acidez fixa, acidez volátil, ácido cítrico, açúcar residual, cloretos, dióxido de enxofre livre, dióxido de enxofre total, densidade, pH, sulfatos e álcool para assim obter a qualidade do vinho em um score de 0 a 10, sendo 10 um vinho excelente vinho e 0 um vinho ruim. Esta base é dividida em vinhos verdes brancos e vinhos verdes tintos.

“*Wine*” a base de dados (*UCI*) classifica vinhos italianos de três diferentes cultivares utilizando 13 parâmetros sendo eles: álcool, ácido málico, cinza, alcalinidade da cinza, magnésio, fenóis totais, flavonoides, fenóis não flavonoides, *Proanthocyanins*, intensidade da cor, matiz, OD280 / OD315 de vinhos diluídos e prolina para assim classificá-los em três classes.

4.1 *Wine Quality White*

Analisaram-se 4898 vinhos verdes brancos nesta base de dados que podem ser classificados de acordo com sua qualidade como vemos na Figura 2.

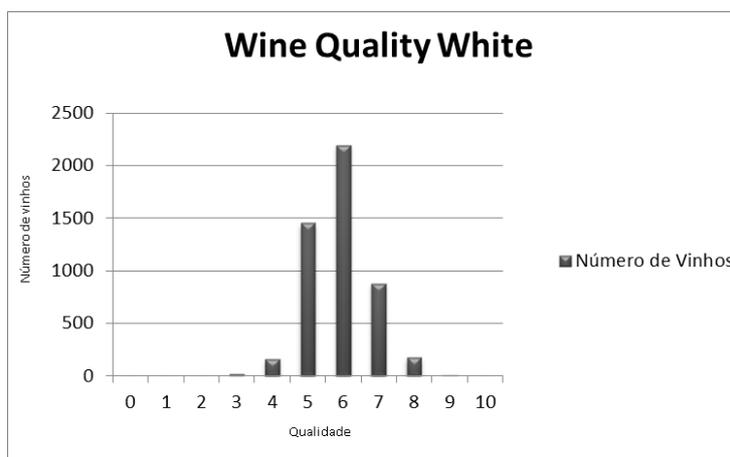


Figura 2 – Número de vinhos por classe do conjunto de dados *Wine Quality White*

Fonte: O autor

Analisando os 11 parâmetros de entrada desta base de dados, conforme tabela 1, poderíamos prever qual seria a qualidade que um vinho verde branco obteria ao ser experimentado por um enólogo.

<i>Wine Quality White</i>	Média	IC95%		Desvio padrão	Moda
		Min	Max		
acidez fixa	6,8548	3,8000	14,2000	0,8439	6,8000
acidez volátil	0,2782	0,0800	1,1000	0,1008	0,2800
ácido cítrico	0,3342	0,0000	1,6600	0,1210	0,3000
açúcar residual	6,3914	0,6000	65,8000	5,0721	1,2000
cloretos	0,0458	0,0090	0,3460	0,0221	0,0440
dióxido de enxofre livre	35,3081	2,0000	289,0000	17,0071	29,0000
dióxido de enxofre total	138,3607	9,0000	440,0000	42,4981	111,0000
densidade	1,5484	0,9871	103,8980	6,8385	0,9920
pH	3,1883	2,7200	3,8200	0,1510	3,1400
sulfatos	0,4898	0,2200	1,0800	0,1141	0,5000
álcool	10,5142	8,0000	14,2000	1,2306	9,5000

Fonte: O autor

Tabela 1 - Estatística descritiva do conjunto de dados *Wine Quality White*

4.2 *Wine Quality Red*

Foram analisados 1599 vinhos verdes tintos nesta base de dados que podem ser classificados de acordo com sua qualidade como vemos na figura 3.

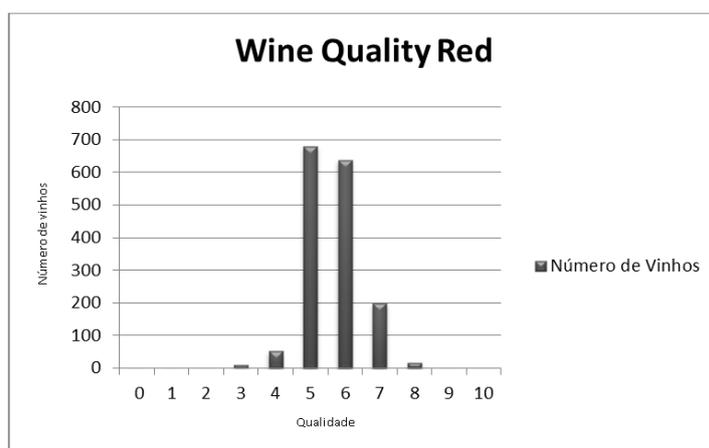


Figura 3 – Número de vinhos por classe do conjunto de dados *Wine Quality Red*

Fonte: O autor

Analisando os 11 parâmetros de entrada desta base de dados (Tabela 2) poderíamos prever qual seria a qualidade que um vinho verde tinto obteria.

<i>WineQualityRed</i>	Média	IC95%		Desvio padrão	Moda
		Min	Max		
acidez fixa	8,3196	4,6000	15,9000	1,7411	7,2000
acidez volátil	0,5278	0,1200	1,5800	0,1791	0,6000
ácido cítrico	0,2710	0,0000	1,0000	0,1948	0,0000
açúcar residual	2,5388	0,9000	15,5000	1,4099	2,0000
cloretos	0,0875	0,0120	0,6110	0,0471	0,8000
dióxido de enxofre livre	15,8749	1,0000	72,0000	10,4602	6,0000
dióxido de enxofre total	46,4678	6,0000	289,0000	32,8953	28,0000
densidade	2,2087	0,9901	100,3690	9,6641	0,9972
pH	3,3111	2,7400	4,0100	0,1544	3,3000
sulfatos	0,6581	0,3300	2,0000	0,1695	0,6000
álcool	10,4751	8,4000	92,3000	2,3081	9,5000

Fonte: O autor

Tabela 2. Estatística descritiva do conjunto de dados *Wine Quality Red*

4.3 Wine

Foram analisados 178 vinhos italianos que podem ser classificados de acordo com a Figura 4 sendo cada uma das três classes um cultivar diferente da Itália.

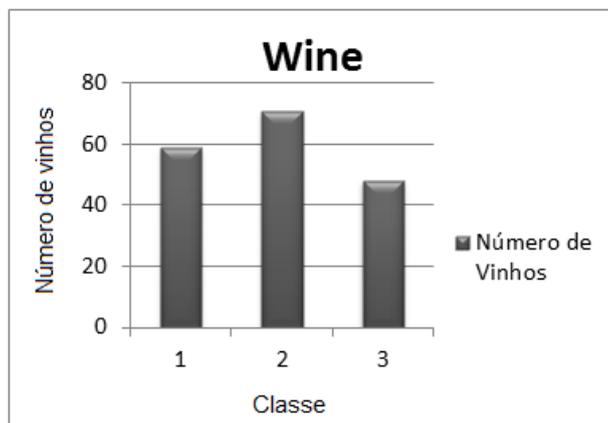


Figura 4 – Número de vinhos por classe do conjunto de dados *Wine*

Fonte: O autor

Analisando os 13 parâmetros desta base de dados para os 178 vinhos criamos a tabela 3 que mostra a distribuição dos parâmetros das amostras.

<i>Wine</i>	Média	IC95%		Desvio padrão	Moda
		Min	Max		
álcool	13,0006	11,0300	14,8300	0,8118	12,3700
ácido málico	2,3363	0,7400	5,8000	1,1171	1,7300
cinza	2,3665	1,3600	3,2300	0,2743	2,2800
alcalinidade da cinza	19,4949	10,6000	90,0000	3,3396	20,0000
magnésio	99,7416	70,0000	162,0000	14,2825	88,0000
fenóis totais	2,2951	0,9800	3,8800	0,6259	2,2000
flavonoides	2,0293	0,3400	5,0800	0,9989	2,6500
fenóis não flavonoides	0,5300	0,1300	0,6600	0,1245	0,2600
proanthocyanins	1,5909	0,4100	3,5800	0,5724	1,3500
intensidade da cor	5,0581	1,2800	13,0000	2,3183	2,6000
matiz	0,9574	0,4800	1,7100	0,2286	1,0400
OD280/OD315	2,6117	1,2700	4,0000	0,7100	2,8700
prolina	746,8933	278,0000	1680,0000	314,9075	520,0000

Fonte: O autor

Tabela 3 - Estatística descritiva do conjunto de dados *Wine*

5. Resultados

Realizando o experimento com o algoritmo *Firework* binarizado, alterando-se a dimensão do problema conforme a base de dados e a quantidade de iterações entre 10, 100 e 1000 pode-se criar a tabela 4 com os resultados para as 3 bases de dados.

			Wine Quality	Wine Quality	Wine
			White	Red	
FWA	10 Iterações	Média das acurácias	58,060%	63,750%	100,000%
		Custo computacional	523,53 s	430,20 s	107,288 s
	100 Iterações	Média das acurácias	56,430%	63,750%	100,000%
		Custo computacional	5186,88 s	3054,11 s	1065,44 s
	1000 Iterações	Média das acurácias	55,890%	63,670%	100,00%
		Custo computacional	11756,08 s	10890,87 s	9367,84 s

Fonte: O autor

Tabela 4 – Acurácias de acerto para as três bases de dados

Os dados da Tabela 7 nos mostram que o algoritmo possui uma acurácia preditiva muito boa por estar acima dos 50% em todos os testes.

Por possuir uma grande variedade de vinhos a base de dados *Wine Quality White* o custo computacional alto em relação as demais bases, porém tal custo pode ser avaliado como, bom visto que além das grandes quantidades de vinhos foram realizadas várias interações com o intuito de obtermos uma análise mais exata dos dados.

Apesar de para as bases de dados *Wine Quality* os resultados não possuírem uma acurácia tão alta quanto para a base *Wine* este método de predição de resultados pode auxiliar muito o produtor, pois ele tem uma previsão de como deve ser a constituição de seu vinho e qual nota ele irá receber aproximadamente ao passar por análise de um especialista.

Comparando a tabela 4 com os resultados obtidos no estudo de Neves Ferreira Gabriel *et al.*, 2016 o algoritmo *FWA* possui acurácia preditiva superior as encontradas utilizando algoritmos de Regressão Logística e Redes Neurais Artificiais (RNA) *Multilayer perceptron (MLP)* independentemente da base de dados ou número de iterações.

6. Considerações Finais

Através do estudo foi possível compreender que reconhecer padrões no agronegócio trás grandes benefícios aos produtores e agricultores de qualquer ramo pois podem antecipar fenômenos que estão por vir. Por meio do *FWA* podemos prever a qual safra o vinho pertence ou qual nota o vinho irá receber se provado por um sommelier e assim o produtor possui confiabilidade que o vinho produzido deverá ter padrões que, teoricamente, garantirão um elevado grau de qualidade.

Estudos sobre vinhos não são muito comuns no Brasil principalmente aqueles que envolvem softwares de inteligência computacional e classificação de dados. O produtor que utilizar o mesmo *software* e algoritmo possivelmente irá melhorar significativamente a qualidade de seus vinhos.

7. Conclusões

Conclui-se que apesar de os resultados para a base de dados Wine Quality serem abaixo de 70% o algoritmo é promissor se comparados com outros algoritmos apresentados no estudo de Neves *et al.* (2016). Portanto o algoritmo atendeu aos objetivos para qual foi proposto podendo, assim, ser utilizado para previsão de dados com confiabilidade.

Referências

AMERINE, A. MAYNARD, PANGBORN, M. ROSE, ROESSLER, B. EDWARD, Principles Sensory Evaluation of Food (Elsevier, USA, 1965).

ARAÚJO, GRACIELI X. Caracterização físico-química de sucos de uva artesanais da região sudoeste do Paraná. 52f. Trabalho de Conclusão de Curso (Bacharelado em Química) da Universidade Tecnológica Federal do Paraná - UTFPR. Pato Branco, 2013.

BRASIL, Lei n° 7678/88 artigo 3°. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/1980-1988/L7678.htm?utm_source=post-producao-do-vinho-como-acontece-a-fermentacao-da-uva&utm_medium=link. Acesso em 10 de janeiro de 2018.

DUTCOSKY, S. D. Análise sensorial de alimentos. 3. ed. Curitiba: Champagnat, 2011.

NEVES, F. GABRIEL, THOM DE SOUZA, R. C., VASQUEZ, C. E., Artificial tasting: computational sensory analysis for wine classification applying logistic regression with bootstrap, International Symposium *International Symposium on Work in Agriculture, 2016*.

SIMON & SCHUSTER, H. *Johnson Vintage: The Story of Wine*, pages 11-16, 1989.

TAN, Y., ZHU, Y., Fireworks algorithm for optimization, in *Advances in Swarm Intelligence* (Springer, Berlin, 2010), pp. 355–364.

THOM DE SOUZA, R. C. Uma metodologia para classificação de dados nominais baseada no processo KDD: ênfase aos algoritmos culturais, estimação de distribuição e análise de correspondência múltipla. Tese (Doutorado em Métodos Numéricos em Engenharia) – Universidade Federal do Paraná - UFPR, Curitiba, Paraná, 2013.