

AGRUPAMENTO DE DADOS COMO METODO DE CLASSIFICAÇÃO DOS CAFÉS NAS REGIÕES DAS MATAS DE MINAS E MANTIQUEIRA

Sammy Guedouani²; Luiza Monteiro Souza³; Aracy Camilla Tardin Pinheiro⁴; Ricardo Rodrigues de Oliveira Neto⁵; Ricardo Henrique Silva Santos⁶; Ney Sussumu Sakiyama⁷; José Luis dos Santos Rufino⁸

¹Trabalho financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

²Graduando em Agronegócio, Departamento de Economia Rural, UFV, Viçosa-MG, sammy.guedouani@ufv.br

³Doutoranda, MSc., Departamento de Fitotecnia, UFV, Viçosa-MG, luiza.m.monteiro@ufv.br

⁴Doutoranda, MSc., Departamento de Fitotecnia, UFV, Viçosa-MG, aracy.pinheiro@ufv.br

⁵Doutorando, MSc., Departamento de Engenharia Florestal, UFV, Viçosa-MG, rick.neto@gmail.com

⁶Professor, DSc., Departamento de Fitotecnia, UFV, Viçosa-MG, rsantos@ufv.br

⁷Professor, DSc., Departamento de Fitotecnia, UFV, Viçosa-MG, sakiyama@ufv.br

⁸Superintendente do Centro de Excelência do Café das Matas de Minas. DSc., Viçosa-MG, rufinojose Luis@gmail.com

RESUMO: Os cafés especiais vêm ganhando grande espaço no mundo, principalmente no agronegócio 4.0 e isso vai até a utilização de sistemas de informação SI e inteligência artificial IA. Uma estatística de análise que vem se destacando é baseada no uso de Redes Neurais Artificiais as RNAs em suma são sistemas computacionais constituídos de entrada, processamento e saída, sistemas similares às do cérebro humano, além do método de plotagem de caixas Box Plots, classificação dos parâmetros. O presente trabalho busca utilizar RNAs nas análises dos resultados da bebida de café para correlacionar as notas das bebidas com alguns elementos agronômicos (modo de produção, cultivar, secagem) e ambientais (altitude) e classificar os elementos ambientais e agronômicos com a qualidade do café (nota final) utilizando a ferramenta de plotagem de caixas Box Plots. Os dados foram obtidos no concurso de cafés especiais organizados pelo SENAR-MG em 2017. Nessas simulações utilizou-se o software STATISTICA 12, tanto para aplicar uma RNA quanto Box Plots. Desse modo foi possível encontrar uma rede neural com índices de treinamento e validação de 88% e 74%, que representa alta correlação entre a variáveis de entrada e saída. Além disso os resultados de dispersão e frequência dos erros percentuais foi inferior a 2,5%, que é interpretado como alta exatidão, e livre de erros. A plotagem de caixas classificou as notas dos cafés com os elementos agronômicos e ambientais, ou seja, cafés do tipo cereja descascado com maiores altitudes e possíveis combinações de cultivar e secador, tinham potencial de produção de cafés com maiores notas. Desse modo, tanto o uso de RNA quanto Box Plots são boas ferramenta de análises para correlacionar a qualidade final de bebida.

PALAVRAS-CHAVE: cafés especiais, redes neurais artificiais, plotagem de caixas, concurso de qualidade.

DATA GROUPING AS A COFFEE CLASSIFICATION METHOD IN THE MINAS AND MANTIQUEIRA REGIONS

ABSTRACT: Specialty coffees have gained a great deal of space in the world, especially in agribusiness. A statistical analysis made by RNAs in short are not systems of use of neural networks, as RNAs in short are computational systems consisting of input, processing and output, systems similar to the human system, besides the box plotting method. parameters. The present work seeks to use RNAs in the fruit plants of mineral acids to correlate as cosmetic notes with some agronomic (mode of production, cultivation, drying) and environmental (altitude) elements and environmental and agronomic plant species with coffee quality (note using a Box Plots box plotting tool. The data were performed in the special event competition organized by SENAR-MG in 2017. In these simulations we used the STATISTICA 12 software, both for the use of RNA and Box Plots. Thus, it was possible to find a neural network with training and validation levels of 88% and 74%, which represents a high frequency between the input and output variables. In addition, the dispersion and frequency of percentage errors was less than 2.5%, which is interpreted as high precision, and free and error free. The plotting of boxes classified as notes of coffees with the agronomic and environmental elements, that is, cherry-type coffees peel at higher altitudes and prefer cultivating and drying combinations, since the potential for producing coffees with higher grades. Thus, both RNA uses, and Box Plots are good analytical tools for correlating final beverage quality.

KEY WORDS: specialty coffees, artificial neural networks, box plots, quality contest.

INTRODUÇÃO

Recentemente, Minas vem mudando o foco para cafeicultura de qualidade, com ênfase no mercado de cafés especiais, ou “*specialty coffee*”. A primeira aparição do termo foi em 1974 no periódico *Tea & Coffee Trade Journal*, cunhado por Erna Knutsen, uma referência mundial no tema de cafés de qualidade. “Ela o usou o termo para descrever como microclimas distintos produzem um grão com um perfil de sabor único” (MARTINEZ, et al, 2015). Acredita-se que o

café terá peculiaridades específicas quanto ao aroma, corpo, sabor, acidez e doçura, dependendo do local e das particularidades de relevo de onde é produzido. Por conta da dimensão territorial do Brasil, é comum um mesmo estado ter diferentes tipos de cafés. A expressão Café Especial se refere ao café classificado com 80 pontos ou mais em uma escala de 0 a 100 pela Metodologia SCA (*Specialty Coffee Association*) de Avaliação Sensorial. Essa classificação contribui em grande parte para estudos relacionadas a caracterização nas regiões cafeeiras. A partir do conjunto de resultados das amostras provadas, são realizadas análises visando caracterizar os cafés de determinada região de origem. Por conta disso, acredita-se que os cafés terão suas peculiaridades devido ao seu terroir, isso significa que elementos ambientais e agronômicos correlacionados entre si, que de certa forma afetam a particularidade específica daquele grão de café. Existem diversos estudos estatísticos que buscam correlacionar os elementos da qualidade do café a um fator seja agronômico ou ambiental, métodos de multivariada e modelos de regressão, (SILVEIRA, 2015). Contudo não existem muitos estudos que correlacionam diversos fatores inter-relacionados para prever a qualidade do café, desse modo surge as inovações em sistema de informação no campo do agronegócio, inteligência artificial, redes neurais artificiais, métodos mais complexos de análise, mas com grande potencial. Redes neurais artificiais (RNA) são sistemas inspirados no cérebro humano, de forma geral constitui de entradas, processamento e saída, as entradas são informações externas carregadas pelo sistema, o processamento corresponde as RNAs, camadas que processam as informações de forma compreenderem os padrões de comportamento e a saída a resposta encontrada pelo sistema, sua interpretação. São compostas por dois sistemas, o primeiro destinado a capacidade de treinamento, aprendizagem (armazena a informação, avalia o padrão de comportamento do banco de dados), e validação (aplica o conhecimento apreendido no treinamento sem a variável resposta). Do ponto de vista prático, resume-se em um sistema computacional paralelo (camadas) constituído de várias unidades de Processamento simples (neurônios artificiais), conectadas entre si de uma maneira específica para executar uma determinada tarefa (BULLINARIA, 2010 apud BINOTI, 2012). Pelo método de RNA os dados inseridos são utilizados para configurar a rede de modo a otimizar os modelos de análise e previsão de resultados. As RNA apresentam modelos de aprendizagem, usando variáveis categóricas (qualitativas) e numéricas (quantitativas), com potencial para avaliação dos cafés. Dentro do modelo de RNA pode-se aplicar outra metodologia a plotagens de caixa (Box Plots). Em Box Plots (este termo foi usado pela primeira vez por Tukey, 1970), faixas ou características de distribuição de valores de uma variável selecionada. Os intervalos de valores de uma variável selecionada (ou variáveis) são plotados separadamente para grupos de casos definidos por valores de uma variável categórica (agrupamento). A tendência central (por exemplo, mediana ou média) e as estatísticas de variação ou variação (por exemplo, quartis, erros padrão ou desvios padrão) são calculadas para cada grupo desse modo as variáveis de entrada, os fatores ambientais e agronômicos. A aplicação da metodologia identifica os elementos e reduz os custos de informação, a partir do modelo é necessário apenas saber quais variáveis produzem determinada qualidade final do café e com isso atender as exigências do mercado internacional, o uso da ferramenta de redes neurais artificiais auxilia o setor do agronegócio que busca cada dia mais pela inovação no campo. Portanto o objetivo desse trabalho busca utilizar RNAs nas análises dos resultados da bebida de café para correlacionar as notas das bebidas com alguns elementos agronômicos (modo de produção, cultivar, secagem) e ambientais (altitude) e classificar os elementos ambientais e agronômicos com a qualidade do café (nota final) utilizando a ferramenta de plotagem de caixas Box Plots.

MATERIAL E MÉTODOS

O projeto foi desenvolvido a partir de dados secundários, já coletados nas regiões Matas de Minas e Mantiqueira de Minas que participaram do último concurso de cafés especiais, organizado pelo SENAR MG no ano de 2017. Nos dados coletados estão disponíveis os laudos contendo informações agronômicas e ambientais bem como as notas sensoriais dos cafés na metodologia SCA. As variáveis de entrada foram: tipo de Processamento, sendo eles cereja descascado ou natural; altitude; cultivares da espécie arábica; modo de secagem do café. Nessa simulação foram selecionadas amostras aleatórias, dispostas por 70% destinadas a treino das redes e 30% para o processo de validação. A princípio foram coletadas 203 amostras de cafés no concurso, contudo pela falta de informação a respeito das variáveis ambientais e agronômicas, algumas amostras foram descartadas. Foram descartadas também amostras incompatíveis com as análises, ou seja, cafés de classificação rio e rio-zona. Desse modo obteve-se um total de 167 amostras. Como as amostras foram avaliadas por 5 provadores, isso resultou em um total de 835 dados prontos para a execução. O estudo desenvolvido neste projeto foi de aproximação das funções, direcionada para a estimação da qualidade das regiões Matas de Minas e Mantiqueira. A proposta consistiu em projetar uma rede neural que aproxime a função desconhecida $f(x)$ e que descreva o mapeamento dos pares de entrada-saída $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ de um conjunto de n padrões de treinamento utilizando redes neurais artificiais, visando otimizar a precisão das estimativas obtidas por meio de regressão e reduzir os erros devido a subjetividade do campo de cafés especiais. As redes treinadas foram do tipo MLP Multilayer perceptron, múltiplas camadas consistindo em duas camadas de neurônios artificiais que processam os dados (camada intermediária e camada de saída). Para obtenção das redes neurais artificiais, foi utilizado o programa STATISTICA 12. De propriedade da STATSOFT, INC. Programa computacional Statistica 12. E.A.U. 2018. As melhores RNA foram selecionadas com base na correlação entre as variáveis agronômicas, ambientais e as notas finais encontradas segundo os avaliadores, estimadas pelas redes (r_{PN}) e na estabilidade dos índices de configuração das redes, fornecidos pelo software nas fases de configuração, seleção e avaliação. Da mesma forma foi verificado quais as

variáveis que mais contribuíram para o desempenho da RNA utilizando a análise de sensibilidade e a análise de plotagens de caixas.

RESULTADOS E DISCUSSÃO

A princípio procurou-se realizar a tarefa de aplicar uma RNA em um modelo de dados qualitativos e quantitativos para estimação da qualidade potencial de um café. Nesse modelo buscou-se estimar através das variáveis se era possível utilizar a ferramenta de redes neurais artificiais como nova abordagem. As configurações para gerar tal modelo foi ajustada de forma a encontrar o melhor valor entre as correlações da nota final entrada (Target) e a nota final saída (Output). Desse modo pode-se encontrar os cinco melhores resultados gerados pelo modelo. Conforme a Tabela 1.

Tabela 1. Características das Redes Neurais selecionadas para correlacionar as variáveis de entrada a nota final a qualidade do café.

RNA	Arquitetura*	Treinamento**	Validação**	Treino. Erro	Validação. Erro***
1	MLP 43-7-1	0,889802	0,746511	1,128770	2,368491
2	MLP 43-10-1	0,889802	0,746511	1,128770	2,368491
3	MLP 43-7-1	0,889802	0,746511	1,128770	2,368491
4	MLP 43-5-1	0,889802	0,746511	1,128770	2,368491
5	MLP 43-4-1	0,889445	0,747371	1,132214	2,368667

* Tipo de rede, número de neurônios (entradas), número de camadas ocultas (Processamento) e número de saída.

**Índices de treino (obtenção da rede) e validação (qualidade da rede), devem ser altos e próximos.

*** Percentagem dos erros percentuais em cada rede.

Observa-se na Tabela 1 que as cinco RNAs obtiveram um índice de treinamento de validação alto e próximos, 88% e 74%. Pode-se dizer que a rede neural apresentou uma alta correlação entre o índice de entradas e saída. O mesmo pode se dizer dos erros, a dispersão e frequência dos erros percentuais apresentam valores inferiores a 2,5%, que segundo Binoti (2012), pode ser interpretado como uma boa exatidão nos resultados.

Encontrado essa correlação entre as variáveis agrônômicas (modo de produção, cultivar, secagem) e ambientais (altitude), podemos afirmar que os elementos acima de certa forma contribuíram para a produção do café. Mas como saber se essa relação foi positiva para o café ou não.

Vamos analisar agora a plotagem de caixas de modo a compreender onde estão localizadas as variáveis ambientais e agrônômicas a nota final do café. Foram realizadas análises de cada variável em relação a nota final, ou seja, para cada fator ambiental e agrônômico existe uma relação com a nota. Desse modo, podemos observar o primeiro gráfico que mostra a relação entre notas dos provadores de cada amostra do café e a classificação das altitudes dessas amostras, conforme a Figura 1.

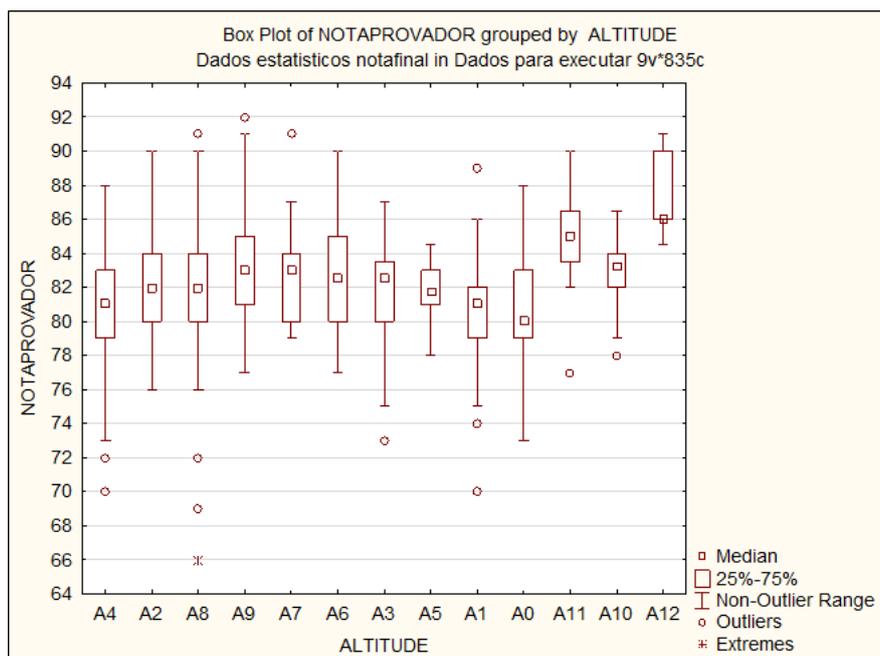


Figura 1. Plotagem de caixas utilizando as notas dos provadores e as classificações de altitudes. ($A1 \leq 800$, $800 < A2 < 850$, ..., $1351 < A12 < 1400$).

Nessa figura podemos observar uma relação entre as altitudes mais elevadas (A10, A11 e A12) de forma crescente e as médias das notas dos cafés, onde cafés produzidos em altitudes mais elevadas apresentaram maiores notas. Contudo, é evidente que mesmo altitudes mais baixas também tiveram apresentações de notas relativamente altas, e uma correlação entre esses dois vetores é observada pela maior concentração em torno das médias. Na figura 2 temos a mesma distribuição utilizando a plotagem de caixas, porem das notas dos provadores em relação as cultivares. As Cultivares contendo abreviação (C) derivam da cultivar catuaí, (MN) mundo novo, (R) Rubi, (CA) catucaí, (B) Bourbon (A) Acaíá.

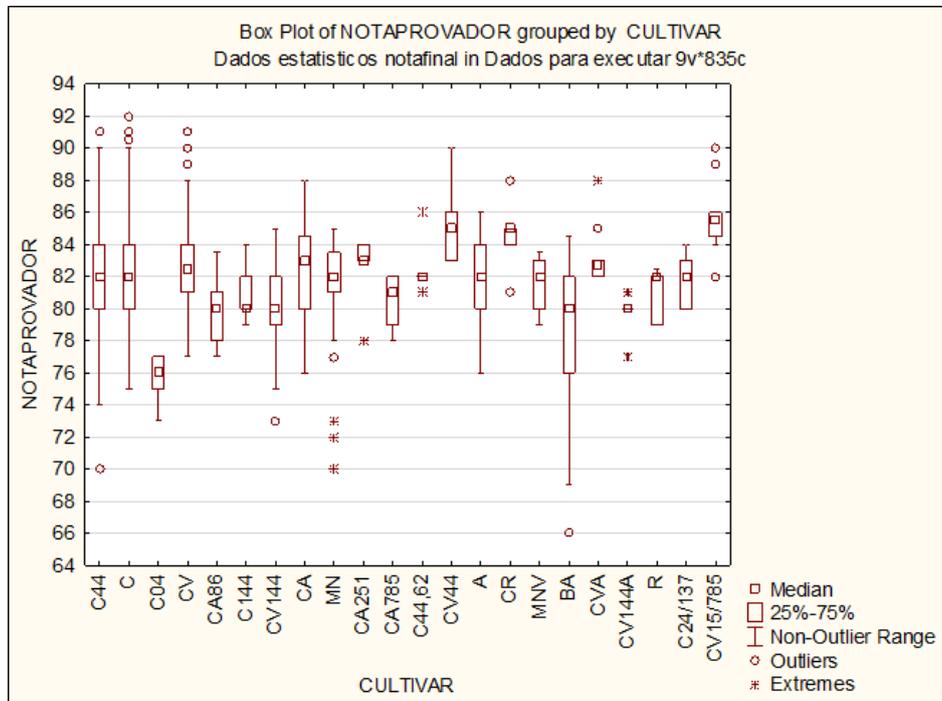


Figura 2. Plotagem de caixas utilizando as notas dos provadores e as classificações das cultivares.

Nessa configuração percebe-se que a distribuição das cultivares em relação as notas dos cafés apresentou maior quantidade de Outliers e Extremos, indicando que cada cultivar pode apresentar diversas classificações de notas, não havendo um padrão de comportamento. Agora vamos analisar a Figura 3, que apresenta a classificação das notas em relação ao método de processamento do grão, seja ele via (N) natural ou (CD) cereja descascado.

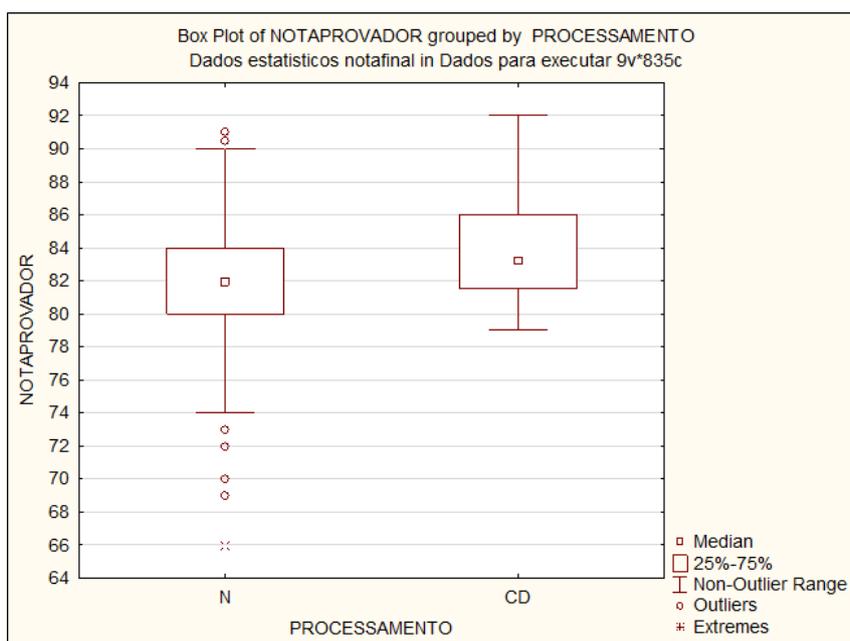


Figura 3. Plotagem de caixas utilizando as notas dos provadores e os métodos de processamento do grão.

É notável na Figura 3 que a concentração de cafés com maiores notas foram cafés via cereja descascado, tanto o agrupamento das médias quanto os Non-Outliers tiveram notas páreas e acima de 80 pontos, indicando bons cafés, já na distribuição do método via (N) natural, apresentou diversos patamares de notas, ou seja, seja acima de 80 pontos quanto abaixo, más com sua nota média mais alta, indicando produção de bons cafés independente dos métodos de processamento

Na Figura 4, temos a distribuição das notas dos provadores em relação ao método de secagem do café, ou seja, (SR) secador rotativo, (TC) terreiro de cimento, (TS) terreiro suspenso, (TA) terreiro asfáltico, (E) estufa, (SC) secador de caixa.

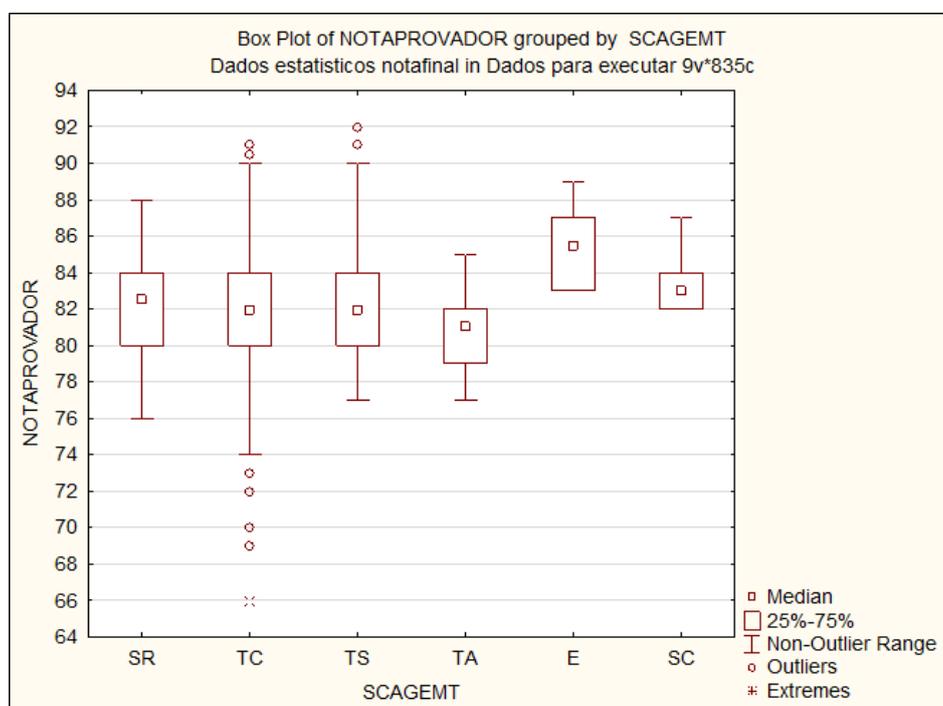


Figura 4. Plotagem de caixas utilizando as notas dos provadores e os métodos de secagem do café.

Na Figura 4, temos que cada método de secagem apresentou uma boa distribuição em relação as notas dos cafés, para o método de secagem (TC), temos pontos extremos em relação a sua média, isso pode ser explicado pelo número de amostras com notas baixas que tiveram esse método na secagem, outro ponto, pode se entender que os demais métodos apresentaram maiores notas indicando que tem melhor eficiência na secagem do café se comparada ao método (TC). Através do método de plotagem caixas tivemos um panorama geral da distribuição dos cafés nas regiões produtoras matas de minas e mantiqueira de minas, mais ainda sobre alguns dos fatores responsáveis pela qualidade desses cafés.

CONCLUSÕES

1. Os resultados utilizando as diversas funções do STATISTICA 12 se mostraram positivos visto que é uma nova abordagem a ser explorada.
2. As correlações entre as variáveis ambientais e agrônômicas apresentaram bons índices de correlação com a nota final do café na rede.
3. O modelo de RNA é uma ferramenta com bons resultados para correlacionar a nota final de bebida de cafés com as variáveis de entrada.
4. O método de plotagem de caixas (BOX PLOTS) nos mostrou como se deu essa distribuição tendo em vista as variáveis e a nota final do café.
5. Portanto se torna valido conhecer bem as variáveis responsáveis para se produzir um bom café, tanto a ferramenta de RNA quanto o BOX PLOTS são instrumentos que nos ajudam nesse processo, ainda mais para um setor de cafés especiais que cresce cada dia mais graças ao avanço da tecnologia.

AGRADECIMENTOS

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de Iniciação Científica Convênio do PIBIC/CNPq/UFV, agradeço também ao Serviço Nacional de Aprendizagem Rural (SENAR) pela disponibilidade dos dados e por fim ao Centro de Excelência do Café das Matas de Minas pelo apoio durante a realização da pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

- BINOTI, MAYRA LUIZA, M.S, Emprego de redes neurais artificiais em mensuração e manejo florestal. 2012. 138f. Tese de Doutorado - Universidade Federal de Viçosa, 2012.
- BSCA BRAZIL SPECIALTY COFFEE ASSOCIATION. Cafés Especiais. 2018. Disponível em: <<http://bsca.com.br/index/home>>. Acesso em: 22 maio.2018.
- MANTIQUEIRA DE MINAS. 2018 Disponível em: <<http://www.mantiqueirademinas.com.br/>>. Acesso em: 22 maio.2018.
- MATAS DE MINAS. 2018. Disponível em: <www.matasdeminas.org.br/>. Acesso em: 22 maio.2018.
- RUFINO, J. Programa Nacional de Pesquisa e Desenvolvimento do Café. Antecedentes, Criação e Evolução. 1º Edição. Embrapa café, Brasília, DF. Embrapa informações tecnológicas, 2006.
- SAKIYAMA, N. S.; MARTINEZ, H, E, P.; TOMAZ, M, A.; BORÉM, A. Café Arábica. do plantio à colheita, Minas gerais, Brasil. Editora UFV, 2015.
- SCAA - SPECIALTY COFFEE ASSOCIATION OF AMERICA. 2018. Disponível em: < <https://sca.coffee/> >. Acesso em: 22 maio.2018.
- SILVEIRA, A. D. (21 de JULIO de 2015). ATRIBUTOS SENSORIAIS DOS CAFÉS CULTIVADOS EM DIFERENTES ALTITUDES E FACES DE EXPOSIÇÃO NA REGIÃO DAS MATAS DE MINAS. p. 60. Tese de Doutorado - Universidade Federal de Viçosa, 2015.
- SILVEIRA, D. D. (20 de outubro de 2014). Estimação do volume de madeira empilhada utilizando imagens digitais e redes neurais. ESTIMAÇÃO DO VOLUME DE MADERIA EMPILHADA UTILIZANDO IMAGENS DIGITAIS E REDES NEURAI, p. 66. Tese de Doutorado - Universidade Federal de Viçosa, 2014.
- STATISTICA 12. De propriedade da STASOFT, INC. Programa computacional Statistica 12. E.A.U. 2018.
- VILELA, P.; RUFINO, J. Caracterização da cafeicultura de montanha de Minas Gerais. Belo Horizonte, 2010.
- ZAIDAN, Ú. R. (31 de JULHO de 2015). QUALIDADE DOS CAFÉS DA 'REGIÃO DAS MATAS DE MINAS' EM FUNÇÃO DA VARIEDADE, ALTITUDE E DA ORIENTAÇÃO DA ENCOSTA DA MONTANHA. p. 43. Tese de Doutorado - Universidade Federal de Viçosa, 2015.