

REDES NEURAIAS ARTIFICIAIS PARA ANÁLISE DA QUALIDADE POTENCIAL DOS CAFÉS ESPECIAIS NAS REGIÕES DAS MATAS DE MINAS E MANTIQUEIRA

Sammy Guedouani²; Aracy Camilla Tardin Pinheiro³; Luiza Monteiro Souza⁴, 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, aracy.pinheiro@ufv.br

⁴Doutoranda, MSc., Departamento de Fitotecnia, UFV, Viçosa-MG, luiza.m.monteiro@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 agronegócio. Suas características peculiares e sofisticação fazem com que fosse apreciado em todo o planeta. No segmento de concurso de qualidade de cafés especiais existe uma metodologia com a qual se atribui uma nota ao café de acordo com suas características básicas. Uma estatística de análise que vem se destacando é baseada no uso de Redes Neurais Artificiais as RNAs são sistemas computacionais constituídos de entrada processamento e saída, organizados de maneira similar às do cérebro humano. O presente trabalho busca utilizar RNAs nas análises dos resultados da bebida de café para associar as notas das bebidas com alguns elementos agronômicos (modo de produção, cultivar, secagem) e ambientais (região, altitude) e prever o potencial qualitativo e quantitativo dos cafés produzidos. Os dados foram obtidos no concurso de cafés especiais organizados pelo SENAR-MG em 2017. Nas primeiras simulações buscou-se provar a possibilidade de uso de RNA, utilizando o software STATISTICA 12, para o qual foram selecionadas as 5 melhores RNAs. Foi possível encontrar uma rede com índices de treinamento e validação de 89% e 75%, que representa alta correlação entre a variáveis de entrada e saída. Além disso os resultados de dispersão e frequência dos erros percentuais foi de 2,5%, que é interpretado como alta exatidão e livre de erros. A análise de Sensibilidade indicou que os parâmetros que mais apresentaram representatividade nas redes foram a cultivar, seguido pela secagem, região, altitude e processamento. Desse modo, o uso de RNA como ferramenta de análises para estimar a qualidade final de bebida se provou uma ferramenta com bons resultados.

PALAVRAS-CHAVE: cafés especiais; redes neurais artificiais; concurso de qualidade.

ARTIFICIAL NEURAL NETWORKS FOR ANALYSIS OF THE POTENTIAL QUALITY OF SPECIAL COFFEE IN THE REGIONS OF WOODS OF MINAS AND MANTIQUEIRA

ABSTRACT: Specialty coffees have been gaining a great deal of space in agribusiness. Its peculiar characteristics and sophistication make it appreciated all over the planet. In the quality competition segment of specialty coffees there is a methodology with which coffee is awarded a grade according to its basic characteristics. One statistic of analysis that stands out is based on the use of Artificial Neural Networks. RNAs are computational systems consisting of input processing and output, organized similarly to those of the human brain. The present work seeks to use RNAs in the analysis of coffee drink results to associate beverage scores with some agronomic (production, cultivation, drying) and environmental (region, altitude) elements and to predict the qualitative and quantitative potential of the coffees produced. Data were obtained from the special coffee competition organized by SENAR-MG in 2017. In the first simulations we tried to prove the possibility of using RNA, using the STATISTICA 12 software, for which the 5 best RNAs were selected. It was possible to find a network with training and validation rates of 89% and 75%, which represents a high correlation between the input and output variables. In addition, the results of dispersion and frequency of percentage errors was 2.5%, which is interpreted as high accuracy and error free. Sensitivity analysis indicated that the most representative parameters in the nets were cultivar, followed by drying, region, altitude and processing. Thus, the use of RNA as an analysis tool to estimate the final drink quality proved to be a good tool.

KEY WORDS: specialty coffees; artificial neural networks; quality contest.

INTRODUÇÃO

O café é a segunda bebida mais consumida no mundo, sendo o Brasil o seu maior produtor e exportador. O destaque vai para o estado de Minas Gerais, responsável por 50% da produção nacional. Adaptando-se a diferentes características edafoclimáticas, a produção do café está localizada em vasta extensão territorial do estado e é dividido em quatro grandes regiões produtoras: Zona da Mata, Sul, Cerrado e Chapada de Minas (VILELA & RUFINO, 2010). Os

Concursos de Qualidades dos Cafés são uma oportunidade de produtores brasileiros mostrarem a alta qualidade dos seus cafés e possibilita a entrada em mercados diferenciados, via leilão ou pela internet, a preços remuneradores. Acredita-se que o café terá peculiaridades específicas quanto ao aroma, corpo, sabor, acidez e doçura, dependendo do local e das particularidades do ambiente onde é produzido. Por conta da dimensão territorial do Brasil, é comum um mesmo estado ter diferentes tipos de cafés. Para Knutsen (1978), grãos dos melhores sabores são produzidos em microclimas especiais (MARTINEZ, et al, 2015). Microclimas encontrados na região das Matas de Minas e Mantiqueira de Minas, regiões estas que foram destaque nos concursos de qualidade dos últimos anos segundo a *brazil specialty coffee association* BSCA. A partir do conjunto de resultados das amostras provadas, são realizadas análises visando caracterizar os cafés de determinada região de origem, com o intuito de reforçar seus atributos de denominação de origem. Por conta disso, acredita-se que os cafés terão suas peculiaridades devido a sua região de origem, 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, métodos de multivariada e modelos de regressão que buscam correlacionar os elementos da qualidade do café a um fator seja agronômico ou ambiental. Contudo não existem muitos estudos que correlacionam os diversos fatores para predição da 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 computacionais inspirados no cérebro humano, de forma geral são constituído por entradas, processamento e saída, as entradas são as informação externas carregadas pelo sistema (variáveis ambientais e agronômica), o processamento corresponde as RNAs, que são camadas que visam coletar as informações e processá-las por meio de um combinador linear de forma compreender os padrões e comportamento da rede, a saída é a nota final a qualidade do café. As RNAs são compostas por dois sistemas: o primeiro destinado à capacidade de treinamento, aprendizagem (armazena a informação, avalia o padrão de comportamento do banco de dados), o segundo à 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. Segundo alguns estudos na área, as RNA apresentam desempenho superior comparado aos modelos de regressão como: estrutura consistente e bem distribuída (camadas); habilidade de aprendizagem e generalização, tornando-se capazes de resolver problemas bem complexos; sua tolerância a erros e falhas ou ruídos; maior capacidade de moldar diversas variáveis e suas relações não-linear; capacidade de adaptar as variáveis categóricas (qualitativas), além das numéricas (quantitativas) (HAYKIN, 2001; BULLINARIA, 2010 apud BINOTI, 2012).

Para se trabalhar com modelos de RNA é necessário saber quais variáveis influenciam determinada qualidade final do café, desse modo as variáveis de entrada, os fatores ambientais e agronômicos, a RNA para processar os dados para prevê-lo, e a variável resposta a qualidade final do café, a nota final. A aplicação da metodologia de RNA's permite reduzir tempo e custos de informação no setor cafeeiro. A partir do modelo é necessário apenas saber quais elementos foram responsáveis para sua produção do café e com isso é possível determinar o potencial de sua qualidade e assim atender as exigências do mercado internacional. Portanto o objetivo desse trabalho é gerar um modelo de predição da qualidade potencial dos cafés especiais usando Redes Neurais Artificiais e associar as características ambientais e agronômicas com a qualidade dos cafés das regiões das Matas de Minas e Mantiqueira para predizer através desse modelo quais variáveis mais influenciam a qualidade final da bebida do café.

MATERIAL E MÉTODOS

O projeto foi desenvolvido a partir de dados secundários, já coletados do 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 com a metodologia SCA. As variáveis de entrada foram: região produzida; tipo de processamento (cereja descascado ou natural); classificação de altitudes; 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 à 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. A partir das escolhas das melhores configurações, elas foram testadas quanto à sua capacidade de prever a qualidade da amostra do café.

RESULTADOS E DISCUSSÃO

A princípio procurou-se realizar a tarefa de provar que é possível a aplicação de 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 foram ajustadas de forma a encontrar o melhor valor entre as correlações da nota final de entrada (Target) e a nota final de 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 estimar a qualidade do café.

RNA	Arquitetura*	Treinamento**	Validação**	Treino. Erro	Validação. Erro***
1	MLP 45-9-1	0,891390	0,747688	1,113443	2,358703
2	MLP 45-11-1	0,891390	0,747688	1,113443	2,358703
3	MLP 45-5-1	0,891390	0,747688	1,113443	2,358703
4	MLP 45-5-1	0,891390	0,747688	1,113443	2,358703
5	MLP 45-10-1	0,891390	0,747688	1,113443	2,358703

* 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, 89% e 75%. 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.

Na Figura 1 podemos ver o gráfico da predição dos valores, X Target (entrada, alvo), Nota final, e Y Output (Saída), Nota final estimada. A Figura 2 mostra o gráfico da dispersão dos erros em relação a nota final Target.

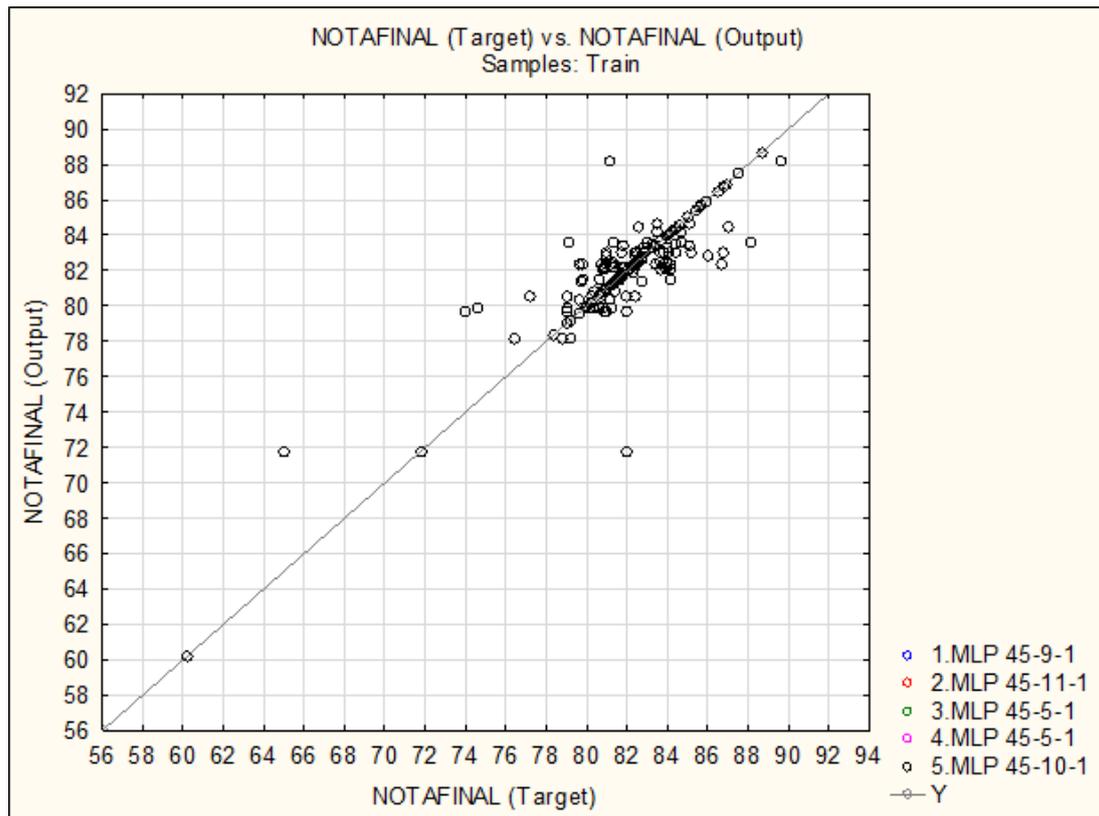


Figura 1. Representa o gráfico das Estimativas das notas finais X Target (entrada) e Y Output (saída).

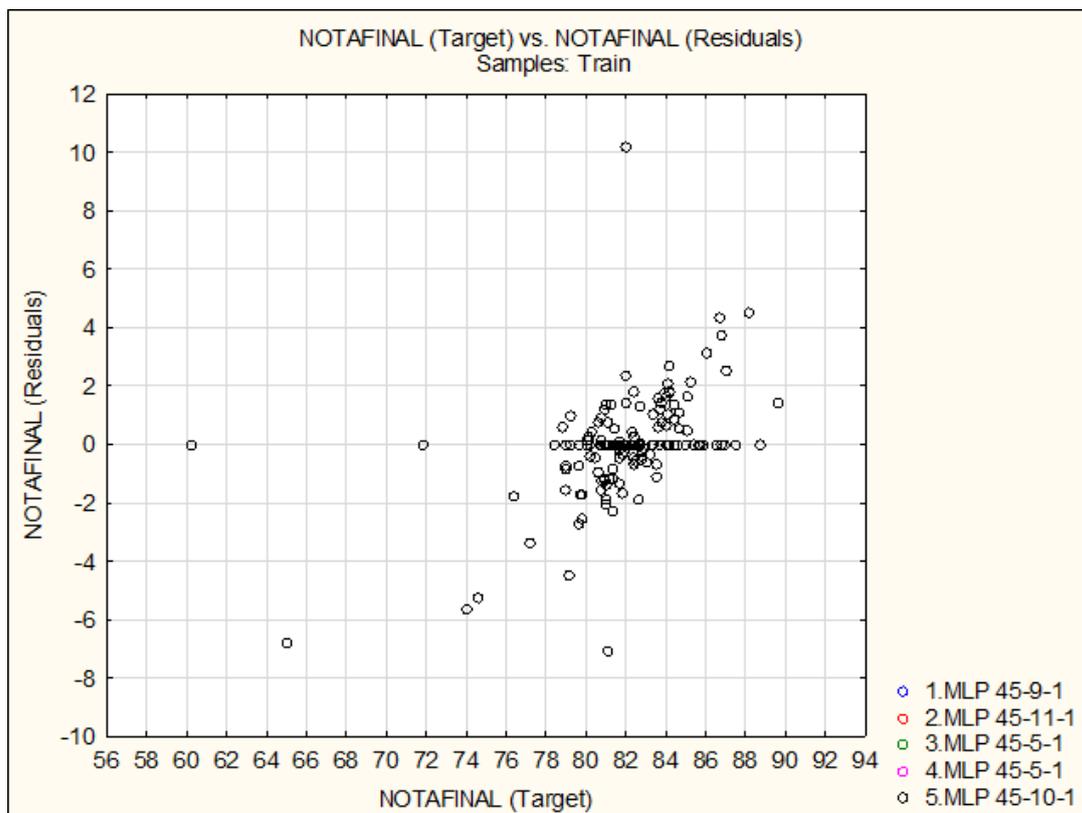


Figura 2. Representa o gráfico da dispersão dos erros percentuais entre nota final X target e a nota final Y Residual.

Como mostrado nas Figuras 1 e 2, as estimativas da RNAs são próximas dos verdadeiros resultados, mostrando a eficiência do uso dessa ferramenta. Embora se encontre algumas amostras fora do eixo, é necessário ressaltar que os dados são notas dos avaliadores de cafés especiais, ou seja, subjetividade, que pode explicar alguns parâmetros encontrados nos resultados. Na Tabela 2 se encontra algumas estatísticas contendo a média, desvio padrão, valor

mínimo e valor máximo para cada variável contínua na análise. Essas estatísticas de dados são divididas por cada amostra (treinamento e validação) e informadas para o conjunto geral de dados.

Tabela 2- Dados estatísticos das notas finais, máximos e mínimos, média, desvio padrão.

Dados Estatísticos	
Mínimo treino	60,20000
Máximo treino	89,60000
Média treino	81,90671
Desvio Padrão Treino	3,29530
Mínimo Validação	60,20000
Máximo Validação	88,12500
Média Validação	81,95930
Desvio Padrão Validação	4,01489
Mínimo Global	60,20000
Máximo Global	89,60000
Média Global	81,92246
Desvio Padrão Global	3,27155

Através das RNAs é possível identificar dentre as variáveis a que mais impactaram no modelo, ou seja, a sua importância ou relevância para a rede neural particular.

A Tabela 3 evidencia que cada variável teve um peso distinto para cada rede neural particular. Esse peso é medido por meio das somas dos quadrados das taxas de resíduos ou de classificação incorreta para o modelo, quando o respectivo preditor for eliminado da rede neural.

Tabela 3. Análise de sensibilidade de cada variável.

Sensitivity analysis (Dados para executar) Samples: Train					
Networks**	CULTIVAR	SECAGEM	REGIÃO	ALTITUDE	PROCESSAMENTO
1.MLP 45-9-1	15,77857	35,86224	49,665730	12,242690	1,000000
2.MLP 45-11-1	15,54132	2,65643	3,163500	5,564670	1,000000
3.MLP 45-5-1	44,30149	12,40459	5,122510	10,999890	1,000000
4.MLP 45-5-1	70,11005	24,82258	7,270140	16,264180	1,000000
5.MLP 45-10-1	35,13155	5,44205	4,540210	8,743740	1,000000
Average*	36,17260	16,23758	13,952420	10,763030	1,000000

* Average (média dos pesos das variáveis); os valores são calculados a partir das somas de quadrados de taxas de resíduos ou de classificação incorreta para o modelo quando o respectivo preditor for eliminado da rede neural. ** Networks, tipo de rede arquitetura da RNA, MLP- Variáveis de entrada. Camadas ocultas. Saída.

Podemos notar que para cada RNA houve um valor relativo ao impacto dessa variável na nota de qualidade do café. Isso indica que para cada uma delas o valor foi sendo ajustado da melhor forma possível, e que as diferentes combinações são possíveis para estimação da nota final. O motivo para a variável cultivar apresentar peso superior em relação as outras é devido a RNA apresentar um maior número de variáveis cultivares, que possibilita as diversas combinações possíveis capazes de alcançar valores superiores. Contudo, não somente isso deve ser atribuído, visto que a variável altitude que possui mais parâmetros em relação a secagem teve menor relevância para a rede, o que significa que a rede não leva somente em consideração a quantidade de parâmetros de cada variável, mas também as possíveis combinações. A rede, portanto, tem uma interpretação simples como uma forma de modelo de entrada-saída, com os pesos como os parâmetros livres (ajustáveis) do modelo.

CONCLUSÕES

Os resultados encontrados utilizando o STATISTICA 12 se mostraram positivos visto que é uma nova abordagem a ser explorada. As variáveis mais importantes para predição da qualidade potencial dos cafés foram cultivar seguidos pela secagem, região e altitude, que possuíam pesos semelhantes e, por fim, a variável processamento, que apresentou a menor interferência na rede. O modelo de RNA é uma ferramenta com bons resultados para estimar a nota final de bebida de cafés. Porém mais estudos relacionados na área devem ser efetuados. Portanto conhecer quais cafés serão

produzidos, e quais variáveis são necessárias para se alcançar esse padrão de qualidade é o ponto chave do processo. Ter essa informação leva a vantagem competitiva no mercado, ainda mais para um setor de cafés especiais que possui alta exigência quanto à qualidade.

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 pôr 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). Estimção do volume de madeira empilhada utilizando imagens digitais e redes neurais. ESTIMAÇÃO DO VOLUME DE MADERIA EMPILHADA UTILIZANDO IMAGENS DIGITAIS E REDES NEURAIAS, 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.