

DISTRIBUIÇÃO ESPACIAL DOS CAFÉS DE MINAS GERAIS POR MEIO DE TÉCNICAS DE MINERAÇÃO DE DADOS

F. P. Bhering; J. N. Barbosa; L. O. A. Junior; P. H. C. Lopes; T. K. B. Silva – Eng. de Controle e Automação CEFET-MG

O Concurso de Qualidade dos Cafés de Minas Gerais é uma iniciativa do Governo do Estado, através da Secretaria de Estado de Agricultura, Pecuária e Abastecimento, coordenado pela Empresa de Assistência Técnica e Extensão Rural de Minas (Emater-MG) e Universidade Federal de Lavras (UFLA). O objetivo do evento é incentivar a constante melhoria da qualidade do café para conquistar mercados, agregar valor ao produto e atender a crescente demanda por produtos diferenciados. A identificação dos melhores cafés de Minas Gerais é realizada pelos provadores da BSCA (*Brazil Specialty Coffee Association*), que avaliam vários aspectos do produto, como cor, aroma, sabor, corpo, acidez, entre outras coisas.

O concurso compreende várias etapas e são avaliados os cafés produzidos nos limites municipais do Estado, sendo avaliados primeiramente os aspectos físicos e, a partir da segunda fase, os atributos sensoriais. A compreensão dos fatores ambientais junto com a inovação das tecnologias podem contribuir com o aumento da produção de cafés especiais em Minas Gerais e a crescente melhoria da qualidade.

Os aspectos climáticos do local de plantio estão intimamente interligados à qualidade de bebida. Em Minas Gerais, em virtude da amplitude das regiões de cultivo é necessário analisar os fatores ambientais (temperatura, altitude, precipitação pluviométrica e umidade) para compreender melhor suas influências na qualidade do café.

O objetivo deste trabalho consiste em associar a qualidade sensorial dos cafés do Concurso de Qualidade - Cafés de Minas com as variações climáticas de cada região do estado por meio de técnicas de mineração de dados, através das informações climáticas, geográficas e classificatórias das amostras de cafés geradas por um banco de dados dos participantes dos concursos nos anos de 2007 e 2008.

Os dados foram disponibilizados em uma planilha eletrônica por meio de uma parceria entre as instituições Universidade Federal de Lavras (UFLA) e Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG). Os registros das notas estavam relacionados aos fatores ambientais das amostras de cada cidade. Somente o atributo cat (categoria da amostra -- Natural ou com Casca) é categórico, o restante são todos numéricos.

A planilha eletrônica foi convertida para uma base de dados relacional -- *MySQL* para facilitar a manipulação dos dados. Foi necessário agrupar os dados por cidades, pois os fatores ambientais coletados representavam somente a cidade das amostras, nesse caso, as 1.161 amostras foram reduzidas em 177 cidades participantes. No agrupamento foi levada em consideração a média aritmética das notas obtidas por cada cidade.

Os agrupamentos foram realizados em três níveis. O primeiro deles foi a média entre as notas de cada etapa do concurso, exceto a primeira etapa que avaliava somente aspectos físicos do café. No segundo agrupamento, calculou-se a média entre as notas das amostras da mesma cidade. E, por fim, uma média entre os dois anos em questão.

Para uma boa análise, há algumas tarefas feitas pela mineração de dados. Neste trabalho, usou-se a tarefa de associação *Apriori* no software *Weka*, com o intuito de identificar associações entre registros de dados que estão ou devem estar relacionados. Sua premissa básica é encontrar elementos que implicam na presença de outros em uma mesma transação.

Foram inseridas no minerador de dados um total de 177 instâncias, divididas em 6 classes. Através da associação de dados utilizando o algoritmo *Apriori*, buscou-se identificar quais atributos mais frequentes estão relacionados à classe nota, através de um suporte mínimo estabelecido de 25% e confiança mínima de 40%.

A mineração de dados mostrou-se uma técnica válida para a associação da qualidade dos cafés correlacionada às características climáticas do ambiente. Ainda assim, para que se alcance um padrão mais próximo da realidade, é fundamental realizar uma nova modelagem com elementos mais coesivos, pois os dados a serem associados foram heterogêneos, prejudicando a associação dos dados e elevando a probabilidade de erros. Logo, um novo estudo com dados mais consistentes já está sendo elaborado para aperfeiçoar o presente trabalho.