

IDENTIFICAÇÃO DE PADRÕES NA PRODUÇÃO DE CANA-DE-AÇÚCAR UTILIZANDO K-MEANS E ANÁLISE MULTIVARIADA

PATRÍCIA F. P. HESPANHOL¹, LUIZ F. S. COLETTA², CAMILA P. C.
GABRIEL⁵, LUIS R. A. GABRIEL FILHO⁶

¹ Graduada e Mestranda em Agronegócio e Desenvolvimento, UNESP FCE, Tupã-SP

² Graduado e Doutor em Ciências da Computação e Matemática Computacional, Professor adjunto, UNESP FCE, Tupã-SP

³ Graduada Mestre e Livre-docente em Matemática, Doutora em Agronomia, Professora adjunto, UNESP FCE, Tupã-SP

⁴ Graduado Mestre e Livre-docente em Matemática, Doutor em Agronomia, Professor adjunto, UNESP FCE, Tupã-SP

Apresentado no
XLVIII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2019
17 a 19 de setembro de 2019 - Campinas - SP, Brasil

RESUMO: O trabalho tem como objetivo identificar padrões na produção de cana de açúcar no estado de São Paulo por meio da utilização de algoritmos de agrupamento de dados. Para isso, adotar-se-á neste trabalho a análise multivariada e o agrupador de dados k-means. Por meio dos dois métodos foi possível identificar os padrões. A temperatura média foi a variável que menos influenciou, enquanto quantidade produzida, área destinada à produção e o índice pluviométrico foram as variáveis que mais influenciaram para a clusterização.

PALAVRAS-CHAVE: Reconhecimento de Padrões, Clusters, Agronegócio.

PATTERN RECOGNITION IN SUGARCANE CULTIVATION USING K-MEANS AND MULTIVARIATE ANALYSIS

ABSTRACT: The objective of this work is to identify patterns in sugarcane production in the state of Sao Paulo, Brazil using data clustering algorithms. For this, we adopted multivariate analysis and the k-means data setter. Through both methods it was possible to identify the patterns. Average temperature was the variable that least influenced, while quantity produced, production area and rainfall index were the variables that most influenced the clustering.

KEYWORDS: Pattern Recognition, Clusters, Agribusiness.

INTRODUÇÃO: A cana de açúcar é uma das principais culturas do agronegócio brasileiro, sendo o país o maior produtor mundial da cultura. São vários os motivos que levam a esse cenário, mas podem-se citar, como principais, as condições climáticas e de solo, que são favoráveis ao cultivo da cana de açúcar e a grande extensão de terras disponíveis (CONAB, 2018; KORB et al., 2016; MIRANDA; VASCONCELOS; LANDELL, 2010; FERNANDES 1984). No Estado de São Paulo, a cana-de-açúcar é o principal produto da agropecuária, com uma produção em 2017 que atingiu 349.200,5 mil toneladas, com uma participação de 35,8% (R\$ 28,07 bilhões) no valor da produção agropecuária e florestal total do estado em 2016 (CONAB, 2018; IEA, 2018).

Foram utilizados, neste trabalho, algoritmos de Aprendizado de Máquina (AM), que é o ramo da Inteligência Artificial que está preocupado com a questão de como construir programas de computador, algoritmos e técnicas que permitam o computador aprender automaticamente com experiências acumuladas para, então, tomar decisões (PEREIRA, 2017; MONARD; BARANAUSKAS, 2003; MITCHELL, 1997) com o objetivo de identificar padrões na cadeia produtiva da cana de açúcar.

Desta forma, procurar-se-á responder a seguinte pergunta: é possível identificar padrões na produção de cana de açúcar no estado de São Paulo por meio de métodos de agrupamento de dados?

Para encontrar os padrões nas características de produção, adotar-se-á neste trabalho métodos de reconhecimento de padrões. O reconhecimento de padrões é a busca automática por descobrir regularidades em dados por meio de algoritmos de computador, sendo que essas regularidades são usadas para realizar ações como classificar dados em diferentes categorias ou agrupá-los (BISHOP, 2006). Para isso, foi utilizada a análise multivariada, que é uma ferramenta que estuda o comportamento de três ou mais variáveis simultaneamente, e o agrupador de dados *k-means*, que é um método de particionamento iterativo, em que o conjunto de dados é particionado criando conglomerados e fazendo várias iterações nesse conjunto (WIVES, 2004).

MATERIAL E MÉTODOS: O trabalho foi conduzido maneira quantitativa, com levantamento de dados secundários realizado por meio de instituições oficiais do Estado de São Paulo. Primeiramente junto à CATI e ao IEA, sendo possível obter informações sobre a área cultivada (hectare), produção de cana de açúcar (tonelada) e produtividade no ano de 2017 em todos os municípios do Estado de São Paulo. Os índices pluviométricos (mm) e de temperatura de cada município foram obtidos por meio do Sistema de Monitoramento Agrometeorológico (Agritempo) disponibilizados pelo Governo Federal. Os dados foram processados no MiniTab e foram elaborados gráficos com o objetivo de obter o agrupamento pelos métodos da análise multivariada e *k-means*.

Para processamento dos dados no *k-means*, foi utilizado o programa de computador *Elki-Data Mining Framework*, em que foi utilizado como documento de entrada a planilha de dados. Como o algoritmo *k-means* possui como pré-requisito o estabelecimento do número de *clusters* do que será obtido com o processamento, foram realizados testes em busca de identificar a melhor quantidade de cluster para a base de dados disponível, por meio do coeficiente de silhueta. O coeficiente de silhueta avalia a coesão dos clusters por meio da proximidade dos pontos do *cluster*. O valor deste coeficiente varia entre $[-1, 1]$, sendo que o valor 1 significa que os dados estão mais próximo do seu *cluster* e -1 representa a que os dados estão mais distantes. Esse coeficiente já está inserido nas configurações do *Elki-Data Mining Framework*. Para processamento da multivariada, foi estabelecida a mesma quantidade de clusters que do método *k-means*, para que fosse possível comparar o resultado dos dois métodos.

Após o processamento dos dados e obtenção dos resultados, foram elaborados gráficos para auxiliar na comparação dos resultados de cada cluster. Além disso, foi elabora um mapa, com o resultado do *k-means*, utilizando o programa de computador Qgis que identifica no estado de São Paulo os *clusters* um e seis. O objetivo deste mapeamento é identificar se há diferença de localização entre os municípios que ficaram inseridos no cluster com maior média produtiva e os municípios do clusters com menor média.

RESULTADOS E DISCUSSÃO: Os resultados obtidos nos métodos *k-means* e multivariada estão demonstrados na Figura 1.

Analisando os gráficos de barras gerados e a relação entre as variáveis, percebe-se que os gráficos de área destinada ao plantio de cana de açúcar e de produção da cultura possuem médias com tendências semelhantes para cada clusters nos dois métodos. Naturalmente, em grandes áreas foram identificados grandes índices de produção, o que inicialmente demonstra que os algoritmos encontraram clusters que condizentes com a realidade.

Considerando o gráfico da produtividade, os clusters com maiores índices de produtividade, são formados por municípios que também possuem grandes áreas e produções. Isso pode significar que a quantidade produzida e área destinada pelos municípios podem influenciar na obtenção de grandes índices de produtividade. No caso da produtividade, o método da multivariada conseguiu identificar rótulos mais distintos entre os clusters, enquanto que o cluster 1obteve uma produtividade maior que 80 nos dois métodos, o cluster 6possui uma

produtividade média menor que 30 hectares por tonelada, enquanto que no k-means essa média é maior que 60 toneladas por hectare, demonstrando que a multivariada foi mais eficiente no agrupamento de municípios somente com produtividade extremamente baixa.

A relação das variáveis produtividade e índices pluviométricos foi inversamente proporcional. Os clusters compostos por municípios com maiores índices de produtividades também são aqueles que obtiveram menor índice pluviométrico no ano de 2017. A cana de açúcar necessita de quantidade de chuva razoavelmente alta, atingindo pelo menos 1200 mm no ano. Ao analisar os clusters formados em relação ao índice pluviométrico, observa-se que o cluster 1, apesar de obter o menor índice pluviométrico, alcançou em 2017 uma quantidade de chuva superior a 1200mm. A boa distribuição das chuvas durante o período de crescimento da cana de açúcar é importante, porém a cultura precisa também de uma época de seca para que ocorra a sua maturação. O clima no Estado de São Paulo favorece esse cenário, pois em geral o Estado tem um período de chuvas bem distribuído (de setembro ou outubro até março ou abril), que coincide com período quente (altas temperaturas) e de crescimento dos colmos da cana.

Essa colocação em relação a necessidade de período quente para crescimento dos colmos da cana vai ao encontro dos resultados obtidos nos clusters para a temperatura média. Pode-se observar que o cluster 1, com maior índice de produtividade, também obteve maior temperatura média no ano de 2017 e o cluster 6, com menor índice de produtividade, também foi aquele com menor temperatura média. Em relação à temperatura média anual a cana é uma planta tropical, portanto um clima mais quente favorece o crescimento da planta.

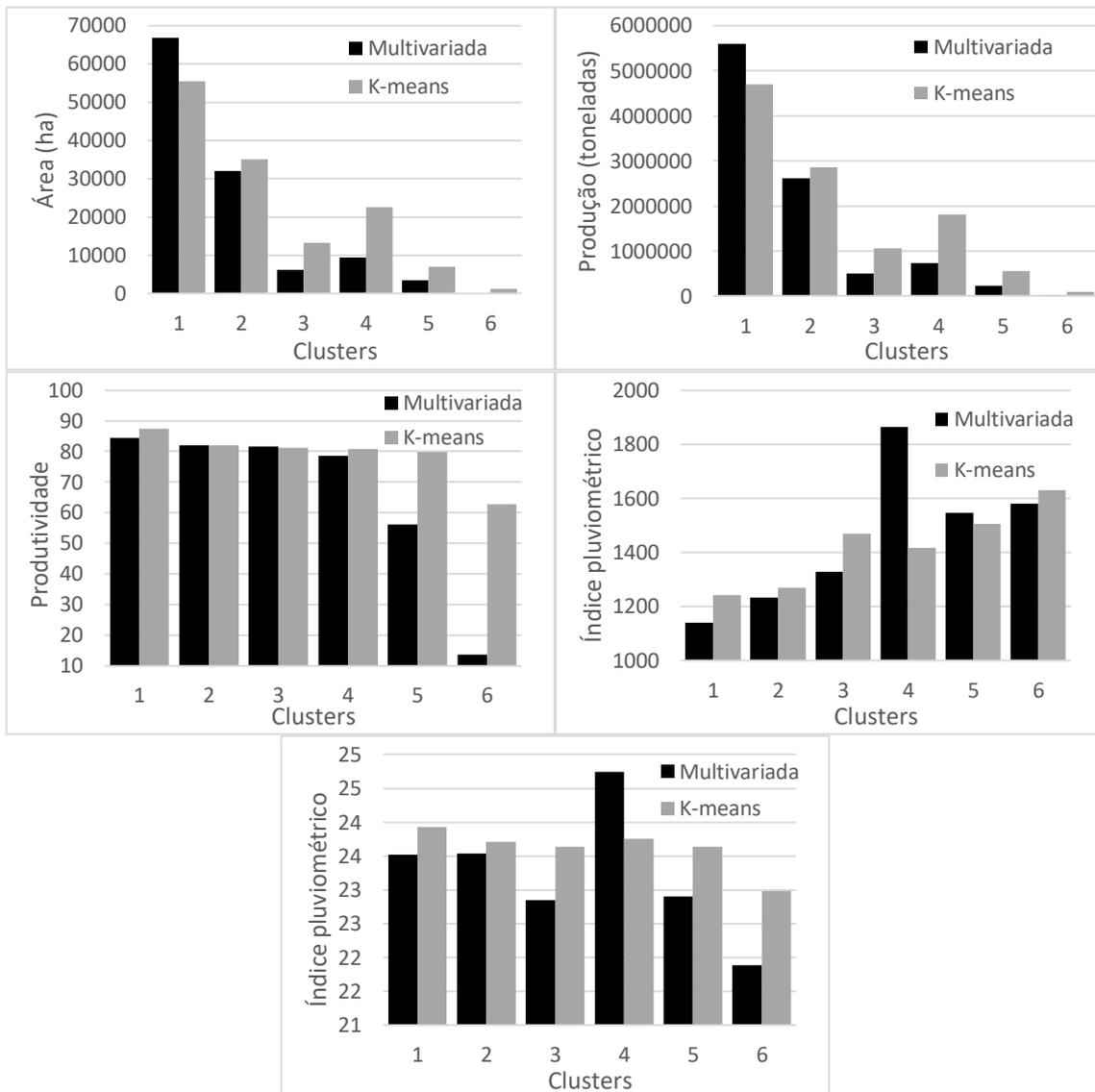


FIGURA 1 – Resultados obtidos por meio do processamento dos dados nos métodos de análise multivariada e k-means.

Os municípios que ficaram inseridos nos clusters com maior média de produtividade foram Barretos, Batatais, Guaíra, Guaraci, Ituverava, Jaboticabal, Morro Agudo, Piracicaba e Valparaíso. A expansão da cana de açúcar no município de Barretos transformou seu espaço geográfico e sua economia, trocando outras atividades agrícolas, como a laranja que era a principal cultura. Em 2017 Barretos destinou uma área de 68 mil hectares para produção de cana de açúcar (IEA, 2019).

Guaraci, devido a decadência da citricultura, expandiu gradativamente a produção de cana de açúcar no município após os anos 2000, atingindo uma área maior que 42 mil hectares em 2017 (IEA, 2019), que foi quando se instalou no município uma usina de cana de açúcar tornando-se a principal atividade econômica e acarretando no crescimento da migração de trabalhadores.

O cultivo e industrialização de cana de açúcar é a principal atividade econômica também de Valparaíso, que possui uma área de aproximadamente 47 mil hectares destinados a cana de açúcar, atingindo em 2017 uma produção de 3,8 milhões de toneladas (IEA, 2019).

Morro Agudo, que é o maior produtor de cana de açúcar do Brasil, com uma área de aproximadamente 96 mil hectares destinados a produção de cana de açúcar, alcançando em 2017 uma produção de aproximadamente 8 milhões de toneladas (IEA, 2019). O município

sempre teve como base econômica a agricultura, porém, as culturas eram soja, milho, sorgo e grãos, porém na década de 1980 produtores rurais começaram a enfrentar grandes dificuldades em manter suas terras expandindo o cultivo de cana de açúcar, que tornou-se a atividade agrícola de maior expressividade principal fonte empregatícia do município (COSTA, CLEPS; 2014). Piracicaba desenvolveu a agricultura do município no decorrer do século XIX com a produção de café e cana de açúcar. No século XX veio o fim do ciclo do café com a estagnação da economia do município e o aumento da industrialização, tornando-se um dos primeiros municípios do país a se industrializar e inaugurar plantas fabris. Mesmo com a mudança da agricultura para o setor de industrialização, Piracicaba ainda possui uma área de aproximadamente 50 mil hectares destinados a cana de açúcar, com uma produtividade média de 90 toneladas por hectare, sendo uma das maiores produtividades do Estado. (IEA, 2019). Para encerrar as discussões envolvendo os resultados do agrupamento obtidos pelo algoritmo *K-means*, o mapa demonstrado na FIGURA 2 representa os municípios que foram agrupados no cluster 1, ou seja, o cluster com maior índice de produtividade (em amarelo) e os municípios agrupados no cluster 6 (em cinza) com menor índice de produtividade.

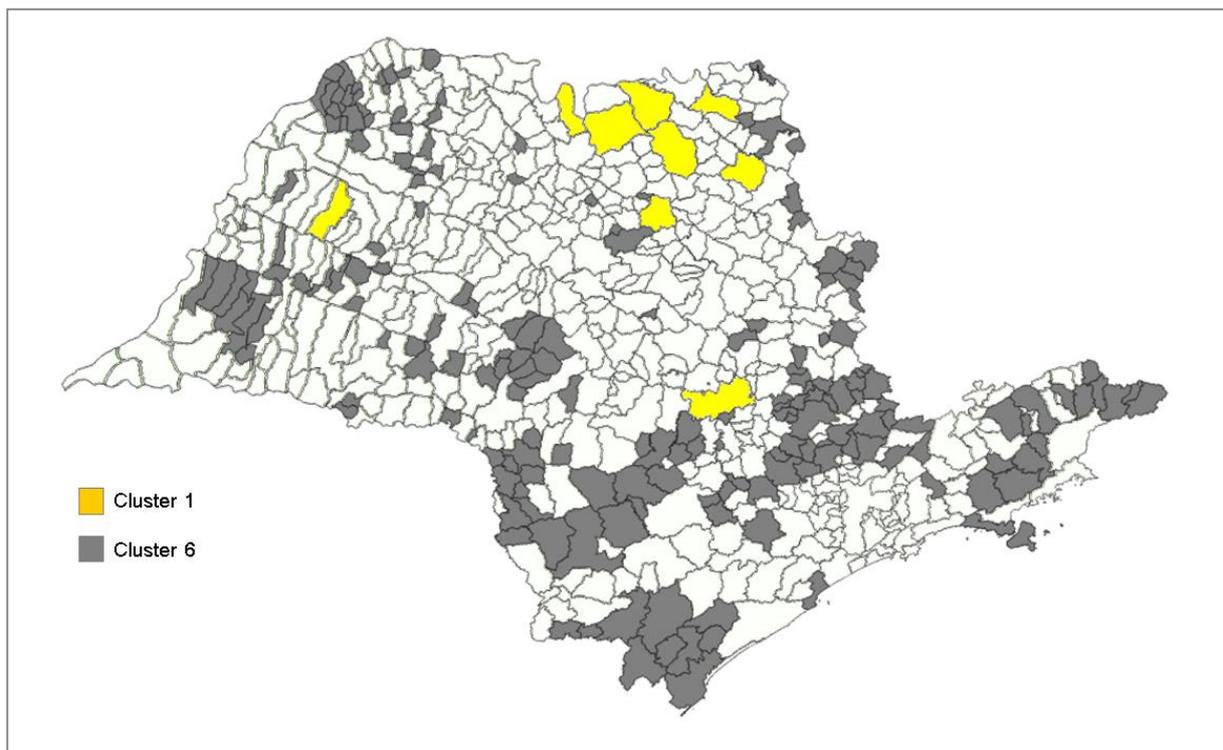


FIGURA 20 – Representação dos municípios pertencentes aos clusters um e seis do K-means.

Percebe-se ainda que a os municípios pertencentes ao cluster um estão concentrados no Nordeste, Oeste e no Centro paulista.

CONCLUSÕES: Os dois métodos analisados forem eficientes no agrupamento de padrões da produção de cana de açúcar no Estado de São Paulo, podendo ser complementares, pois cada um deles permite uma análise mais aprofundada em determinados aspectos. Entende-se que a temperatura média foi a variável que menos influenciou no agrupamento dos dados, enquanto que a quantidade produzida, área destinada a produção de cana de açúcar e o índice pluviométrico tiveram maior influência sobre o agrupamento. No mapa elaborado com os resultados do agrupamento dos algoritmos foi possível identificar que os municípios analisados

que possuem como padrão maior média de produtividade estão concentrados na parte superior do mapa, não sendo identificados municípios com esse padrão no sul do estado de São Paulo.

REFERÊNCIAS:

BISHOP, C. M. **Pattern Recognition and Machine Learning**. New York: Springer, 2006.

CONAB. Companhia Nacional de Abastecimento. **Série histórica das safras. 2018b**. Disponível em: <<https://www.conab.gov.br/index.php/info-agro/safras/serie-historica-das-safras?start=20>>. Acesso em: 10 mar. 2019.

COSTA, A. L. S.; CLEPS, G D G. A produção sucroalcooleira em Morro Agudo (SP) e a migração piauiense. **CAMPO-TERRITÓRIO: revista de geografia agrária**, v. 9, n. 17. 2014.

FERNANDES, A. J. **Manual da cana de açúcar**. Livroceres: Piracicaba, 1984.

IEA. Instituto de Economia Agrícola.. **Cana de açúcar**. 2018a. Disponível em: <<http://ciagri.iea.sp.gov.br/nia1/cadeia/cadeiaCana.aspx>>. Acesso em: 05 fev. 2018.

IEA. Instituto de Economia Agrícola. **Estatísticas da Produção Paulista**. 2019. Disponível em: < http://ciagri.iea.sp.gov.br/nia1/subjetiva.aspx?cod_sis=1&idioma=1>. Acesso em: 06 fev. 2019.

KORB, V.; ROSA, C. P.; GUERRA, D.; SANTOS, J. S.; SOUZA, E. L. Produtividade de cultivares de cana-de-açúcar de ciclo precoce na região noroeste do rio grande do sul. **Salão do Conhecimento**, v. 2, n. 2, 2016.

MIRANDA, L. L.; VASCONCELOS, A. C. M. D.; LANDEL, M. G. A. **Cana de açúcar**. Campinas: Instituto Agrônômico de Campinas, 2010.

MITCHELL, T. M. **Machine learning**. New York: WCB, 1997.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre Aprendizado de Máquina. In: REZENDE, S. O. **Sistemas Inteligentes - Fundamentos e Aplicações**. Barueri: Manole Ltda, 2003. p. 89-114.

PEREIRA, T. S. **Uso de inteligência artificial para estimativa da capacidade de suporte de carga do solo**. 2017. 179 f. Tese (Doutorado em Engenharia Agrícola) - Universidade Federal de Santa Maria, Santa Maria, 2017.

WIVES, L. K. **Utilizando conceitos como descritores de textos para o processo de identificação de conglomerados (clustering) de documentos**. 2004.136 f. Tese (Doutorado em Computação) - Universidade Federal do Rio Grande do Sul, Porto Alegre – RS. 2004.