

Modelos Preditivos para Classificação de Aptidão Agrícola de Municípios

Predictive Models for Classification of Agricultural Aptitude of Municipalities

Carolina Lobello Lorensini¹
Stanley Robson de Medeiros Oliveira²
Daniel de Castro Victoria³

Resumo – Este trabalho teve como objetivo desenvolver e validar modelos baseados em técnicas de mineração de dados para classificar municípios com aptidão agrícola, na região do Matopiba. Foram utilizados dados socioeconômicos e físicos dos municípios, obtidos do Instituto Brasileiro de Geografia e Estatística (IBGE). Os modelos preditivos foram gerados por meio de indução de árvores de decisão juntamente com métodos de seleção de atributos. Os resultados encontrados revelaram o potencial das técnicas de mineração de dados para a classificação de áreas com aptidão agrícola.

Termos para indexação: mineração de dados, seleção de atributos, Matopiba, sustentabilidade.

Abstract – This work aimed to develop and validate models based on data mining techniques to classify municipalities with agricultural aptitude in the MATOPIBA region. Socioeconomic and physical data of the municipalities were used, obtained from IBGE, The predictive models were generated through induction of decision trees along with attribute selection methods. The results showed the potential of data mining techniques for the classification of areas with agricultural aptitude.

Index terms: data mining, attribute selection, Matopiba, sustainability.

1 Engenheira ambiental e sanitária, mestranda em Engenharia Agrícola da Faculdade de Engenharia Agrícola da Universidade Estadual de Campinas (Feagril/Unicamp), Campinas-SP

2 Bacharel em Ciência da Computação, Ph.D. em Ciência da Computação, pesquisador da Embrapa Informática Agropecuária, Campinas-SP

3 Engenheiro Agrônomo, Doutor em Ciências, pesquisador da Embrapa Informática Agropecuária, Campinas-SP

Introdução

A classificação da aptidão agrícola das terras, baseada no nível tecnológico, apoia a gestão territorial do agronegócio por relacionar o recurso natural solo com a produtividade, buscando a conservação e a sustentabilidade agroambiental. Mapas de aptidão agrícola fornecem informações objetivas que podem ser aplicadas no planejamento agrícola e na avaliação do uso das terras, permitindo apontar áreas com uso adequado, subutilizadas ou que suportariam a intensificação do uso (Ramalho Filho; Beek, 1995; Valladares et al., 2007).

Para que ocorra o desenvolvimento de políticas agrícolas no Brasil, é necessário conhecimento sobre as potencialidades de uso do território e seus recursos naturais, bem como o mapeamento do uso atual das terras. Assim, é possível avaliar a adequabilidade do uso atual e avaliar alternativas para o desenvolvimento futuro.

Em particular, a região denominada Matopiba, ganhou importância no setor agropecuário brasileiro. A região apresenta forte crescimento desde os anos 2000 e atualmente é considerada fronteira em expansão e intensificação do Brasil (Guerra, et al., 2017). Números do IBGE mostram o desenvolvimento agrícola na região. De acordo com a Produção Agrícola Municipal (PAM) (IBGE, 2018), a área plantada de algodão, milho e soja, na região do Matopiba era de 21.470.086 ha em 2003. No ano 2010, a área plantada era 27.486.222ha e, em 2016, aumentou para 39.140.994 ha, um crescimento em 82,30%.

A topografia plana, os solos profundos e o clima favorável ao cultivo das principais culturas de grãos e fibras possibilitaram o crescimento vertiginoso da região (Borghi et al., 2014).

A proximidade de terminais portuários, terras comparativamente de baixo valor comercial, possibilidade de escoamento da produção por meio de outros modais de transporte – alternativos ou complementares ao modal rodoviário – e perspectiva de diminuição de custos de produção em função de melhorias de capacidade logística a serem implantadas nos portos do Arco Norte tornam a região promissora e atrativa do ponto de vista do agronegócio (Belchior et al., 2017). Diante desse contexto, pode-se dizer que existe um potencial agrícola no Matopiba, mas é preciso aprimorar pesquisas para identificar e mapear os recursos disponíveis e a capacidade da região. A avaliação do potencial da terra constitui um estágio muito importante nos estudos ambientais para fins de zoneamento e planejamento. A avaliação da aptidão agrícola é um processo interpretativo e pode sofrer alterações com avanços tecnológicos, por isso deve ser atualizada.

Este trabalho tem como objetivo desenvolver e validar modelos baseados em técnicas de mineração de dados para classificar municípios com aptidão agrícola, na região do Matopiba.

Material e Métodos

O estudo foi realizado na região do Matopiba que é dividido entre os quatro estados: 33% no Maranhão (135 municípios); 38% no Tocantins (139 municípios); 11% no Piauí (13 municípios) e 18% na Bahia (30 municípios), como mostra a Figura 1. A topografia plana, os solos profundos e o clima favorável ao cultivo das principais culturas de grãos e fibras possibilitaram o crescimento vertiginoso da região (Borghi et al., 2014).

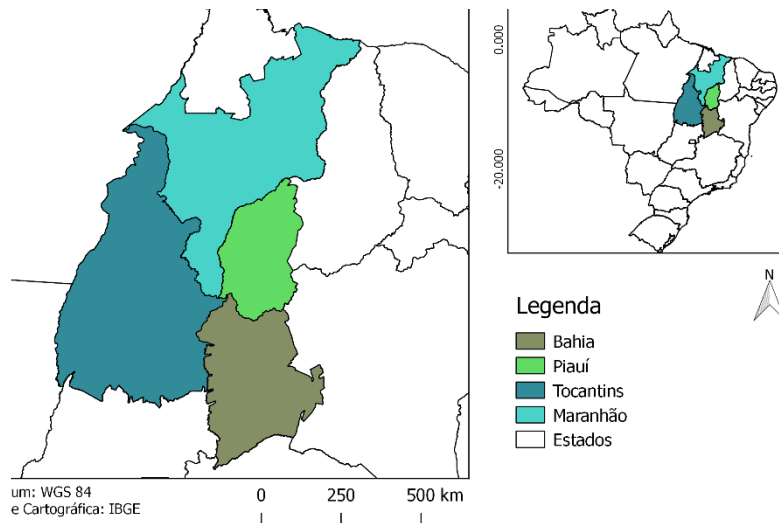


Figura 1. Localização da área de estudo.

Para construir modelos de classificação a fim de reconhecer áreas com aptidão agrícola, na região do Matopiba, utilizou-se dados socioeconômicos e físicos dos municípios. Os dados socioeconômicos são: área plantada, quantidade produzida e rendimento médio, no período de 10 anos, entre 2006 até 2016 das culturas de soja, algodão e milho (1ª e 2ª safras). Esses estão disponíveis no banco de dados da Produção Agrícola Municipal (PAM), Produto Interno Bruto (PIB), população, Índice de Desenvolvimento Humano Municipal (IDHM), taxa de alfabetização, renda per capita, fluxo migratório interno e externo dos municípios. Todos os dados foram obtidos do IBGE. Os dados de caracterização física incluem: precipitação, umidade relativa e temperatura para o mesmo período, declividade, elevação e tipo de solo. Os dados do clima (precipitação, temperatura e umidade relativa) foram obtidos a partir de estações meteorológicas. A declividade e a elevação foram obtidas do produto Shuttle Radar Topography Mission (SRTM) com resolução de 30 metros. Para o solo, o IBGE disponibilizou o mapa de pedologia com a escala cartográfica de 1:250.000 que permite um nível de detalhamento regional (IBGE, 2018).

O atributo-meta (ou variável resposta) foi definido a partir da classificação de aptidão agrícola feita por Lumbreras et al. (2015), seguindo a metodologia de Ramalho Filho e Beek (1995). Para realizar esse trabalho foram utilizadas as classes de aptidão agrícola: boa e regular para os três níveis de manejo (A, B e C) e com a utilização de lavoura. Essas classes representaram áreas aptas para agricultura. As classes restrita e inapta, para a utilização de lavoura, pastagem plantada, silvicultura e pastagem natural, representam áreas não aptas para a agricultura e, assim, foi possível definir a classe de cada município da região do Matopiba.

A quantidade de municípios que formaram o banco de dados para o estudo foi de 327 (representando 327 instâncias), onde 196 representaram a classe “não apta”, 131 a classe “apta” e dez municípios foram retirados do banco de dados, pois sua área total está dentro de reserva legal.

A classificação ocorreu com a aplicação do algoritmo J48 (Árvore de decisão), também conhecido como C4.5 (Quinlan, 1993), disponível no ambiente de software WEKA 3.8.1 (Witten; Frank, 2005). O algoritmo C4.5 baseia-se em uma medida de separação de dados derivada da “entropia” que representa o ganho de informação, aplicada sucessivas vezes aos dados, onde cada nível da árvore representa um atributo que particiona melhor o subconjunto, segundo essa medida.

Para seleção de atributos foram utilizados os seguintes métodos (Han et al., 2011):

Correlationbased feature selection (CFS), Ganho de Informação (InfoGain), Wrapper com

algoritmo J48 e Qui-quadrado (X^2).

Os modelos gerados foram avaliados utilizando o método de validação cruzada (cross-validation) com 10-folds, além das métricas: a) acurácia, b) índice Kappa; e c) precisão por classe.

Resultados e Discussão

Os algoritmos de seleção de atributos foram utilizados para eleger aqueles com melhor poder preditivo para as duas classes. A Tabela 1 mostra os atributos selecionados pelos métodos em estudo. O CFS, InfoGain e o Qui-quadrado destacaram os atributos de área plantada, rendimento médio, quantidade produzida das culturas de soja e milho. E também os atributos físicos de temperatura média e mínima, precipitação e solo. O Wrapper, executado conjuntamente com o algoritmo J48, selecionou um número menor de atributos. No entanto, também destacou os atributos: umidade relativa mínima (2009) e renda per capita (2000).

O Wrapper geralmente alcança melhores resultados por ser executado com algoritmo indução e seus dados de treinamento (Hall, 1999). O método selecionou subconjuntos de atributos baseado no algoritmo indicado, J48, e avaliou a qualidade de cada subconjunto com o mesmo algoritmo que foi utilizado na classificação. Com isso, encontrou o melhor conjunto de atributos que se adequa ao algoritmo. O InfoGain e o Qui-quadrado são métodos em que cada atributo é avaliado com relação à classe (atributo-meta). No final, foi feita uma classificação que privilegia os atributos com menor entropia e maior valor do qui-quadrado para os métodos InfoGain e o Qui-quadrado, respectivamente.

Tabela 1. Atributos escolhidos pelos métodos de seleção de atributos.

Atributo	Método			
	CFS	Wrapper	InfoGain	Qui-quadrado (X^2)
AP-2008 (Milho 1ª Safra)				
AP-2011 (Milho 1ª Safra)				
AP-2012 (Milho 1ª Safra)				
AP-2013 (Milho 1ª Safra)				
AP-2014 (Milho 1ª Safra)				
AP-2015 (Milho 1ª Safra)				
AP-2014 (Soja)				
AP-2015 (Soja)				
Precipitação 2015				
Precipitação 2008				
Precipitação 2007				
Precipitação 2005				
QP-2008 (Milho 1ª Safra)				
QP-2010 (Milho 1ª Safra)				
QP-2012 (Milho 1ª Safra)				
QP-2014 (Milho 1ª Safra)				
QP-2014 (Soja)				
QP-2015 (Soja)				
QP-2016 (Soja)				
Renda per Capita (2000)				
RM-2006 (Milho 1ª Safra)				
RM-2011 (Milho 1ª Safra)				
Solo				
Temp. Máxima 2006				
Temp. Máxima 2008				
Temp. Máxima 2010				
Temp. Máxima 2014				
Temp. Máxima 2013				
Temp. Média 2006				
Temp. Média 2007				
Temp. Média 2009				
Temp. Média 2011				
Temp. Mínima 2005				
Temp. Mínima 2006				
Temp. Mínima 2007				
Temp. Mínima 2008				
Temp. Mínima 2012				
Temp. Mínima 2013				
Umidade Relativa 2009				

A Tabela 2 apresenta os resultados das métricas para os classificadores com e sem a seleção de atributos. Na comparação dos valores de acurácia, índice Kappa e precisão por classe, o método de seleção de atributos que apresentou o melhor resultado foi o Wrapper com a acurácia de 84.4 % e com índice Kappa de 0.67. Segundo Landis e Koch (1977), o desempenho do classificador é considerado bom. A precisão da classe “não apta” é de 0.86 e da classe “apta” é 0.81. Para este caso, em particular, foram geradas 31 regras a partir da árvore de decisão.

Os métodos Qui-quadrado e InfoGain obtiveram valores semelhantes com relação à acurácia (79,81%) e ao índice Kappa (0,58). A precisão da classe “não apta” foi de 0.83 e o número de regras foi 35, em ambos os casos. A única diferença foi na precisão da classe “apta”, isto é, 0,74 e 0,76, para os métodos Qui-quadrado e InfoGain, respectivamente. Por outro lado, o algoritmo J48, sem seleção de atributos, gerou um modelo preditivo com performance inferior aos demais. Sem a seleção de atributos, o banco de dados pode conter atributos redundantes e irrelevantes que reduzem a exatidão da classificação e a qualidade dos grupos encontrados, já que atributos redundantes ou irrelevantes não agregam valor ao modelo gerado (Tan et al., 2009)

Tabela 2. Resultados dos modelos gerados por meio do algoritmo J48 com diferentes métodos de seleção de atributos

	Acurácia	Kappa	Precisão “Não Apta”	Precisão “Apta”	Número de regras
Sem Seleção	78.28	0.54	0.79	0.75	37
X²	79,81	0.58	0.83	0.74	35
InfoGain	79.81	0.58	0.83	0.76	35
Wrapper	84.40	0.67	0.86	0.81	31
CFS	74.92	0.48	0.80	0.67	31

Conclusões

1. Os resultados encontrados indicam o potencial das técnicas de mineração de dados para a classificação de áreas com aptidão agrícola ou não, utilizando dados disponíveis no Produção Agrícola Municipal (PAM-IBGE) e dados físicos.

2. Métodos de seleção de atributos contribuíram para melhorar os resultados dos modelos de classificação gerados, sendo que o método Wrapper com algoritmo J48 obteve o melhor resultado.

Agradecimentos

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela bolsa concedida - Código de Financiamento 1691359; à Faculdade de Engenharia Agrícola – Feagri/Unicamp e à Embrapa Informática Agropecuária, pelo suporte.

Referências

BELCHIOR, E. B.; ALCANTARA, P. H. R.; BARBOSA, C. F. **Perspectivas e desafios para a região do Matopiba**. Palmas: Embrapa Pesca e Aquicultura, 2017. 3 p. (Embrapa Pesca e Aquicultura. Fronteira agrícola, 16). Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/156581/1/CNPASA-2017-fa16.pdf>>. Acesso em: 20 set. 2018.

BORGHI, E.; BORTOLON, L.; AVANZI, J. C.; BORTOLON, E. S. O.; UMMUS, M. E.; GONTIJO NETO, M. M.; COSTA, R. V. da. Desafios das novas fronteiras agrícolas de produção de milho e sorgo no Brasil: desafios da região do MATOPIBA. In: KARAM, D.; MAGALHÃES, P. C. (Ed.). **Eficiência nas cadeias produtivas e o abastecimento global**. Sete Lagoas: Associação Brasileira de Milho e Sorgo, 2014. cap. 25, p. 263-278.

GUERRA, B. J.; SCHULTZ, B.; SANCHES, I. D. Mapeamento automático da expansão da agricultura anual no MATOPIBA entre 2002 e 2015 utilizando a plataforma Google Earth Engine. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 18., 2017, Santos. **Anais...** São José dos Campos: INPE, 2017.

HALL, M. **Correlation based feature selection for machine learning**. 1999. 198 f. Thesis (Degree of Doctor of Philosophy) - University of Waikato.

HAN, J.; KAMBER, M.; PEI, J. **Data mining**: concepts and techniques. 3rd ed. Burlington: Elsevier, 2011. 703 p.

IBGE. Sistema IBGE de Recuperação Automática. **Produção agrícola municipal**. 2018. Disponível em: <<https://sidra.ibge.gov.br/pesquisa/pam/tabelas>>. Acesso em: 20 set. 2018.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159-174, 1977.

LUMBRERAS, J. F.; CARVALHO FILHO, A. de; MOTTA, P. E. F. da; BARROS, A. H. C.; AGLIO, M. L. D.; DART, R. de O.; SILVEIRA, H. L. F. da; QUARTAROLI, C. F.; ALMEIDA, R. E. M. de; FREITAS, P. L. de. **Aptidão agrícola das terras do Matopiba**. Rio de Janeiro: Embrapa Solos, 2015. 48 p. il. color. (Embrapa Solos. Documentos, 179).

QUINLAN, J. R. **C4.5**: Programs for Machine Learning. San Mateo: Morgan Kaufmann Publishers. 1993. 302 p. il. (Morgan Kaufmann series in machine learning).

RAMALHO FILHO, A.; BEEK, K. J. **Sistema de avaliação da aptidão agrícola das terras**. 3. ed. Rio de Janeiro: EMBRAPA-CNPQ, 1995. 65 p.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introdução ao data mining**. Rio de Janeiro: Ciência Moderna, 2009. 896 p.

VALLADARES, G. S.; QUARTAROLI, C. F.; HOTT, M. C.; MIRANDA, E. E. de; NUNES, R. da S.; KLEPKER, D.; LIMA, G. P. **Mapeamento da aptidão agrícola das terras do Estado do Maranhão**. Campinas: Embrapa Monitoramento por Satélite, 2007. 25 p.

WITTEN, I. H.; FRANK, E. **Data mining**: practical machine learning tools and techniques. 2nd ed. Amsterdam; Boston: Morgan Kaufmann, 2005. 525 p. il. (Morgan Kaufmann series in data management systems).