



MINERAÇÃO DE DADOS PARA IDENTIFICAÇÃO DE FLORESTAS PLANTADAS DESTINADAS À PRODUÇÃO DE BIOENERGIA UTILIZANDO IMAGENS DE SATÉLITE

Carlos Tavares Nonato¹ & Yolanda Vieira de Abreu²

RESUMO: Neste trabalho, avaliou-se a acurácia da classificação e identificação de áreas cultivadas com florestas plantadas com finalidade energética em imagens orbitais do sensor Landsat5 TM, por meio de técnicas estatísticas de mineração de dados. Os *pixels* foram convertidos em valores de refletância de superfície, nas vizinhanças dos municípios de São Miguel do Tocantins, São Bento do Tocantins, Araguatins, Babaçulândia, Darcinópolis e Wanderlândia, na região Norte do Estado do Tocantins. Foram utilizados atributos de textura para identificar melhorias nos resultados da classificação. As técnicas de mineração de dados utilizadas se mostraram eficientes na identificação precisa de florestas plantadas em imagens do satélite Landsat 5, tanto pelo desempenho da classificação, quanto pela redução da quantidade de informação necessária para realizar a identificação. Os modelos de árvore de decisão, por meio do algoritmo J48, alcançaram taxas de acerto superiores a 90% na identificação de espécies plantadas em meio a outros alvos. Assim, as técnicas empregadas neste estudo possibilitaram o desenvolvimento de modelos de classificação robustos no auxílio ao planejamento e à tomada de decisão sobre o plantio de florestas no território brasileiro.

PALAVRAS-CHAVE: Florestas plantadas, agroenergia, seleção de atributos, árvores de decisão, classificação de imagens.

IDENTIFICATION OF PLANTED FORESTS DESTINED TO BIOENERGY PRODUCTION THROUGH DATA MINING USING SATELLITE IMAGES

ABSTRACT: In this work, we evaluated the accuracy of the classification and identification of areas cultivated with planted forests by satellite images of Landsat 5 TM, through data mining statistics techniques statistics. The *pixels* were converted to surface reflectance values from the cities of Sao Miguel do Tocantins, Sao Bento do Tocantins, Araguatins, Babaçulândia, Darcinopolis and Wanderlandia, located in northern state of Tocantins. Texture attributes were used to identify improvements in the classification results. The data mining techniques used were effective in the accurate identification of planted forests in Landsat 5 satellite images, both by the classification performance and by reducing the amount of information needed to perform identification. The tree models, through the J48 algorithm, achieved success rates above 90% in the identification of species planted among other targets. Thus, the techniques used in this study enabled the development of robust classification models to assist the planning and decision-making on forest plantations in Brazil.

KEYWORDS: Planted Forests, agro-energy, feature selection, decision trees, and pictures classification.

1 INTRODUÇÃO

A energia derivada da biomassa cultivada, também chamada de agroenergia, tem se destacado como fonte renovável devido à tendência mundial de priorizar tecnologias que proporcionem sustentabilidade ambiental, melhor qualidade e maior segurança no fornecimento de energia.

Os investimentos em florestas plantadas têm se intensificado a partir do desenvolvimento de tecnologias que possibilitam sua utilização em aplicações bioenergéticas, como por exemplo, a área de produção de bio-óleo, gás de síntese, hidrogênio e etanol, além dos usos clássicos na siderurgia, lenha e carvão vegetal. (DAMASO *et al.*, 2013).

¹ Mestre em Agroenergia, Universidade Federal do Tocantins (UFT).
E-mail: crltavaresnonato@gmail.com

² Prof^a. Dr^a. Mestrado em Agroenergia - Universidade Federal do Tocantins (UFT). E-mail: yolanda@uft.edu.br

Segundo Leite (2012) pode-se notar uma mudança na geografia da cadeia produtiva agroflorestal atual onde culturas como o Eucalipto e a Seringueira, que contam

com décadas de pesquisas realizadas nas regiões Sul e Sudeste do Brasil e apoio de grandes instituições de pesquisas do país, estão migrando para novas regiões de fronteiras no território brasileiro. O estado do Tocantins, na região Norte do Brasil, está entre as novas áreas de fronteira agrícola.

Desdobram-se, então, novos desafios técnico-científicos para a adequação dos conhecimentos já consolidados às novas regiões para que se possa alcançar a sustentabilidade econômica e ambiental das culturas.

Nesse sentido, as geotecnologias têm sido amplamente utilizadas para responder as demandas relativas ao planejamento bioenergético (FIORESE e GUARISO, 2010), pois permitem que múltiplas e diferentes variáveis possam ser correlacionadas espacialmente e temporalmente.

Segundo Ramachandra (2009), na área da agricultura, as aplicações de sensoriamento remoto relacionam-se às atividades de levantamento e caracterização dos solos, estimativas de produtividade agrícola e identificação e mapeamento das culturas.

A disponibilidade de grande quantidade de informação nas imagens de sensoriamento remoto e o desenvolvimento da área de reconhecimento de padrões tem permitido que técnicas de mineração de dados, como por exemplo, os algoritmos estatísticos, redes neurais e as árvores de decisão, se apresentem como alternativa promissora na resolução de problemas de identificação e classificação das diferentes coberturas do solo (CELINSKI, 2008).

O Plano Estadual de Florestas (PEF/TO) da região Norte do estado do Tocantins prevê triplicar a área com plantios florestais em um horizonte de 25 anos. Esta decisão vincula-se a necessidade do aumento da produção de madeira para geração de energia nas áreas de celulose, siderurgia, carvão, lenha e de látex, acompanhando a crescente demanda destes setores industriais. Por isso optou-se por utilizar o Estado do Tocantins como estudo de caso deste trabalho. Utilizou-se da técnica de mineração de dados e a árvore de decisão que foi avaliada na classificação de *pixels* puros de áreas com florestas plantadas em meio a alvos diversos, como áreas urbanas, solo exposto, corpos d'água, e outros tipos de vegetação.

O objetivo foi avaliar a acurácia da classificação e identificação de áreas cultivadas com florestas plantadas para fins energéticos em imagens do sensor remoto TM a bordo do satélite Landsat 5.

2 MATERIAL E MÉTODOS

2.1 Base de dados com reflorestamentos

A base de dados com a distribuição dos reflorestamentos no Estado do Tocantins, no período de 2011/2012 foi obtida junto a Agência de Defesa Agropecuária do Estado do Tocantins (ADAPEC, 2013). As variáveis

disponíveis para cada propriedade eram: nome do município, nome da propriedade, latitude, longitude, área em hectares, espécie plantada e variedade.

2.2 Áreas de estudo

Para seleção da área de estudo, municípios do Estado do Tocantins, utilizou-se como critério a concentração de florestas plantadas na região geográfica bem como regiões com segmentação da produção voltada à obtenção de energia (SEMADES-TO, 2013). Os municípios selecionados e as respectivas coordenadas geográficas (Datum SAD69) são apresentados na Tabela 1.

Tabela 1 - Áreas de estudo.

Municípios incluídos	Coordenadas Geográficas
São Miguel do Tocantins São Bento do Tocantins	Lat 04° 52' 51,10'' a 06° 41' 26,21'' Lon 47° 57' 59,76'' a 46° 35' 21,12''
Araguatins	Lat 04° 52' 49,40'' a 06° 41' 24,07'' Lon 49° 30' 09,36'' a 48° 08' 00,96''
Babaçulândia Darcinópolis Wanderlândia	Lat 06° 19' 34,39'' a 08° 08' 10,39'' Lon 48° 16' 42,24'' a 46° 53' 47,40''

Fonte: O autor.

Utilizou-se, também, imagens obtidas junto ao Catálogo de Imagens do Instituto Nacional de Pesquisas Espaciais (INPE). Devido à grande quantidade de imagens disponíveis, foram consideradas imagens que apresentaram o menor percentual de cobertura de nuvem e datas de imageamento correspondentes ao período de corte dos plantios. Na Tabela 2 são apresentadas as informações relativas às órbitas/ponto, horários de passagem do satélite e ângulo de incidência do sol.

Tabela 2 - Cenas Landsat 5 TM utilizadas no trabalho.

Órbita Ponto	Data e hora do imageamento	Ângulo em graus
222/64	17/09/2011 às 13h06	64,24°
222/65	14/09/2010 às 13h29	59,45°
223/64	07/08/2011 às 13h12	51,89°

Fonte: O autor.

2.3 Seleção das amostras de cobertura do solo

Para o treinamento do classificador, foram utilizados cinco tipos de coberturas de solo conforme ilustra a Figura 1.

As áreas circundadas com a cor azul representam regiões contendo rios e lagos. As áreas circundadas com a cor roxo representam regiões com reflorestamentos. As áreas

circundadas com a cor amarela representam solo exposto. Áreas circundadas com a cor verde representam

vegetação natural e por fim áreas circundadas na cor vermelha representam regiões contendo área urbana.

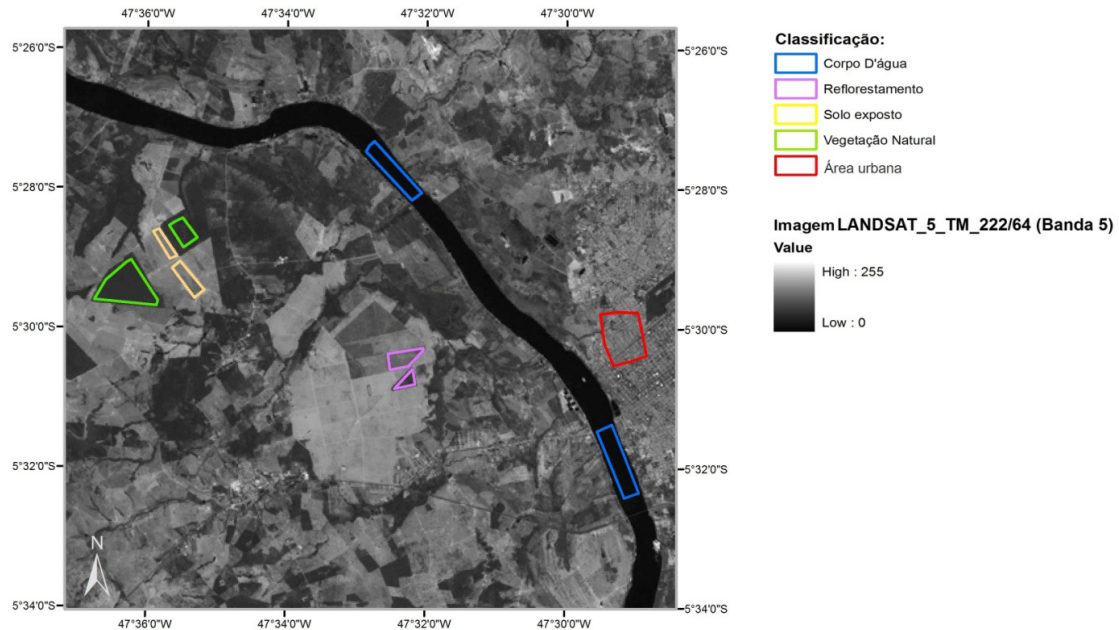


Figura 1 - Seleção de amostras de pixels relativas às classes de cobertura a serem classificadas: em azul, regiões contendo rios e lagos; em roxo, reflorestamentos; em amarelo, solo exposto; em verde vegetação natural; por fim na cor vermelha, regiões contendo área urbana.

2.4 Conversão de Números Digitais (NDs) para valores de reflectância

O sensor TM capta a intensidade de radiância proveniente da área do terreno e registra em *pixels* de imagens digitais. Este valor numérico é chamado de nível de cinza (ND) e seu valor pode variar entre 0 a 255 (8 bits), dependendo da quantidade de bits utilizada para cada *pixel*.

Para a realização da conversão para valores de reflectância utilizou-se o método proposto por Chander e Markhan (2003), usando os coeficientes de calibração listados na Tabela 3. Este procedimento de calibração foi aplicado em cada uma das imagens utilizadas. A equação utilizada para a conversão é apresentada a seguir:

$$L = L_{min} + \frac{ND * (L_{max} - L_{min})}{255} \quad (1)$$

Onde:

L é o valor de radiância calculado a partir do número digital relativo ao *pixel*; *ND* é número digital do *pixel* da imagem variando entre 0 e 255; *L_{max}* é o valor máximo do coeficiente de pós-calibração da banda; *L_{min}* é o valor mínimo do coeficiente de pós-calibração da banda.

Tabela 3 - Pós-calibração sensor Landsat 5/TM.

Banda	L _{min}	L _{max}	Ganho	Offset
1	-1.52	193.0	0.762824	-1.52
2	-2.84	365.0	1.442510	-2.84
3	-1.17	264.0	1.039880	-1.17
4	-1.51	221.0	0.872588	-1.51
5	-0.37	30.2	0.119882	-0.37
6	1.2378	15.303	0.055158	1.2378
7	-0.15	16.5	0.065294	-0.15

Fonte: Chander e Markhan (2003).

2.5 Índices de vegetação

A fim de ampliar e enriquecer o espaço de atributos-base para o treinamento do modelo de classificação, foram utilizados cinco diferentes índices de vegetação baseados em combinações lineares e não lineares das bandas espectrais do sensor TM (TANAJURA et al., 2005). Estes índices são:

Tabela 4 - Fórmulas dos índices de vegetação e referências.

Índice	Fórmula
NDVI	$(B4 - B3) / (B4 + B3)$
PVI	$((B3 - B3_{solo})^2 + (B4 - B4_{solo})^2)^{1/2}$
SAVI	$((1 + L) * (B4_{solo} - B3_{solo})) / (B4_{solo} + B3_{solo} + L)$
RVI	$B4 / B3$
EVI	$G * (B3 - B4) / (L + B4 + (C1 * B3) - (C2 * B1))$

Fonte: Ponzoni e Shimabukuro (2007).

Na Tabela 4, os parâmetros B3 e B4 são respectivamente as bandas 3 e 4 do sensor Thematic Mapper do satélite Landsat 5 e correspondem às faixas espectrais do azul, vermelho e infravermelho próximo respectivamente; B3_{solo} e B4_{solo} são as médias dos valores dos *pixels* dos solos expostos para as bandas 3 e 4 respectivamente e L é uma constante igual a 0,5.

2.6 Atributos de Textura

Embora a informação fornecida pelos sensores remotos envolva essencialmente elementos espectrais, essa não é a única informação que pode ser extraída destes dados.

As texturas contêm informações importantes sobre o arranjo estrutural das superfícies e suas relações com o entorno e Haralick *et al.*, (1973) propõe uma metodologia para explorar a dependência espacial dos níveis de cinza da imagem para montar a matriz que melhor representa a variação tonal denominada “matriz de co-ocorrência”, que expressa a quantidade de combinações de diferentes níveis de cinza ocorrem em uma imagem.

Na construção da matriz considera-se a direção e a distância (θ , d) entre *pixels* próximos, conforme ilustrado

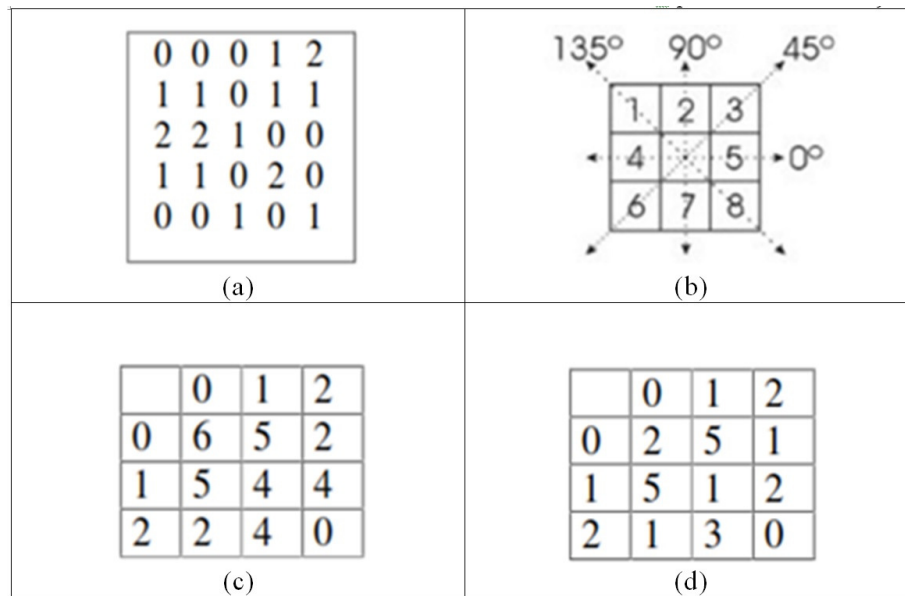


Figura 3 - Ilustração da construção da matriz de co-ocorrência: (a) Matriz ilustrando uma janela 5x5 com níveis de cinza (0, 1, 2); (b) Ângulos utilizados no cálculo da matriz de co-ocorrência; (c) Matriz de co-ocorrência com um pixel de vizinhança na direção 45°; (d) Matriz de co-ocorrência com um pixel de vizinhança na direção 90°.

Fonte: Vasconcelos (2010).

2.7 Mineração de dados

Segundo Fayyad (1996), KDD (Knowledge Discovery in Databases) refere-se ao processo global de descoberta do conhecimento a partir de bases de dados. Este, também pode ser entendido como uma intersecção de áreas de pesquisa afins como aprendizado de máquina, reconhecimento de padrões, estatística, visualização de dados e computação, sendo a atividade de mineração uma das fases principais deste processo.

Dentre as principais técnicas utilizadas em mineração de dados, destacam-se as técnicas baseadas em crescimento poda e validação como as árvores de decisão (HAN e KANBER, 2006 *apud* NONATO, 2010).

Em uma árvore de decisão tem-se que cada nó interno representa um teste em um atributo preditivo, uma ramificação partindo de um nó interno representa um resultado para o teste, uma folha da árvore representa um rótulo de classe e uma nova observação é classificada seguindo um caminho na árvore da raiz até a folha.

na Figura 3 que apresenta a matriz de uma imagem com três níveis de cinza, os ângulos frequentemente utilizados no cálculo da matriz de co-ocorrência e as matrizes resultantes da aplicação do cálculo nas direções 45° e 90° com 1 *pixels* na vizinhança, respectivamente.

Haralick (1973) *apud* Batista e Haertel (2006) sugere que os principais atributos para descrever a textura a partir da matriz de co-ocorrência são: média dos *pixels* vizinhos, variância dos *pixels* vizinhos, homogeneidade, contraste, entropia, segundo momento e correlação.

O algoritmo de indução utilizado neste trabalho foi o J48, que faz parte da família de algoritmos de indução mais conhecida na área de mineração de dados, desenvolvido por Quinlan (1993) *apud* Lenz (2009).

Foram testadas diferentes configurações de atributos, onde cada uma delas resultou em diferentes taxas de acerto do modelo, bem como tamanhos diferentes de árvores.

2.8 Avaliação da classificação

Segundo Han e Kamber (2006), uma maneira para solucionar o problema de *overfitting*, que tem como consequência uma alta taxa de erro na aplicação do modelo em novos dados, consiste em dividir aleatoriamente o conjunto de dados em dois subconjuntos independentes, geralmente dois terços para o conjunto de treinamento e um terço para o conjunto de teste.

Após definidos o conjunto de treinamento e o de testes, o próximo passo da avaliação é a aplicação do modelo ao conjunto de testes selecionado. Como resultado o

analista obtém a chamada matriz de confusão representada na Tabela 5 e amplamente utilizada em análises estatísticas de concordância.

Tabela 5 - Matriz de confusão de dimensão 2x2.

		Predita	
		C ₊	C ₋
Verdadeira	C ₊	VP	FN
	C ₋	FP	VN

Fonte: Nonato (2010).

Para explicar a intensidade da concordância entre dois ou mais julgadores, ou entre dois métodos de classificação, utilizamos a medida *Kappa* que é baseada no número de respostas concordantes, ou seja, no número de casos cujo resultado é o mesmo entre os julgadores. O *Kappa* é uma medida de concordância e indica como cada classificação difere de uma classificação aleatória dos tipos de cobertura. O coeficiente *Kappa* é definido pela equação (WITTEN *et al.*, 2011):

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (2)$$

Onde,

- $Pr(a)$ é a concordância relativa observada para uma dada classe na matriz de confusão;
- $Pr(e)$ é a probabilidade de concordância esperada para esta mesma classe.

O cálculo do coeficiente *Kappa* leva em consideração todas as classes em estudo.

Tabela 6 - Resultados da classificação para diferentes conjuntos de atributos.

Método	Melhores atributos selecionados (Por ordem de mérito)	Taxa de Acerto do Modelo	Estatística KAPPA	Número de Regras
Ganho de informação e Qui-quadrado	SAVI, PVI, NDVI, RVI, B7, EVI	86,92%	0,83	84
Razão de ganho de informação	EVI, NDVI, RVI, B7, SAVI, B5	92,90%	0,90	52
CfsSubsetEval	B7, B5, B2, NDVI, EVI, B1_Media	92,83%	0,90	61
Somente as Bandas	(B1, B2, B3, B4, B5 e B7)	93,81%	0,92	57
Bandas e índices de vegetação	(B1, B2, B3, B4, B5, B7, NDVI, RVI, PVI, SAVI)	93,59%	0,91	58
Todos os atributos		93,59%	0,92	59

Fonte: O autor.

2.9 Conjunto de dados utilizado no trabalho

O conjunto de dados resultante do processo de preparação de dados foi composto por 60 atributos (59 atributos preditores e um atributo resposta). O atributo resposta refere-se as classes de cobertura do solo e os atributos preditores foram compostos pelas seis bandas do espectro do visível, cinco índices de vegetação e 48 atributos de textura.

A cota de dados utilizados para treinamento do modelo foi de 66% e os 34% restantes foram utilizados para testes com todos os métodos aplicados.

3 RESULTADOS

A Tabela 6 refere-se a quatro métodos de seleção de atributos, considerando a contribuição individual de cada atributo e os resultados considerando subconjuntos de atributos frequentemente utilizados na área de sensoriamento remoto.

Verificou-se que o melhor conjunto de atributos obtido é composto por 6 dos 60 atributos analisados e correspondentes as bandas do espectro. Nota-se também que os atributos de textura não foram relevantes na melhoria do processo de classificação, pois não aparecem em nenhum dos subconjuntos selecionados pelos métodos aplicados.

Os métodos de seleção apresentaram boa concordância ao selecionar subconjuntos bastante semelhantes, o que sugere que o subconjunto selecionado pode ser considerado uma boa escolha.

Ao todo foram utilizados 52.847 registros, dos quais 34.876 (66%) foram utilizados para treinamento e 17.971 para testar a classificação em todas as diferentes configurações de atributos como mostra a Tabela 6.

A Tabela 7 apresenta os resultados da classificação dos diferentes alvos por meio da matriz de confusão. A maior taxa de falsos positivos foi observada na distinção entre Solo exposto e Área urbana e entre Solo exposto e Florestas plantadas.

O maior número de ocorrência de erros de classificação entre Solo exposto e Área urbana sugere que a classificação feita pelo modelo de árvore de decisão é coerente, dado que estes atributos são relativamente semelhantes. Já a ocorrência de classificações de Solo exposto pode ter como causa a dificuldade do algoritmo

classificador em diferenciar as amostras dos plantios iniciais ou com corte recente de áreas com solo exposto. As medidas relativas à qualidade do modelo estão listadas na Tabela 8. Verifica-se que a precisão na maioria das classes foi maior que 90%.

A classe com menor desempenho da classificação foi solo exposto, que como mencionado anteriormente teve o maior número de casos confundidos com área urbana. Nota-se também que a classificação do atributo Corpos d'água, isto é, rios e lagos teve precisão de 100%.

Tabela 7 - Matriz de confusão para o melhor modelo de classificação.

Classe	Classificado como				
	Área urbana	Vegetação Natural	Florestas plantadas	Solo exposto	Corpos d'água
Área urbana	4349	1	116	328	0
Vegetação Natural	6	1730	106	0	0
Florestas plantadas	63	77	5182	91	0
Solo exposto	160	0	162	2756	0
Corpos d'água	1	0	0	0	2840

Fonte: O autor.

A poda em modelos de árvore de decisão é um ajuste com o objetivo de evitar *overfitting* e também considerado um meio de deixar o modelo de classificação mais compreensível, já que ocorre a diminuição da árvore de decisão. Os níveis de pré-poda

são equivalentes aos números mínimos de *pixels* da folha. Na Tabela 9 são apresentadas a Acurácia ou Taxa de acerto, a estatística Kappa e o número de regras geradas para níveis diferentes de pré-poda.

Tabela 8 - Métricas de qualidade do modelo extraídas da matriz de confusão.

Classe	Sensitividade	Confiabilidade Positiva (precisão)	Curva ROC	F-measure
Área urbana	0,907	0,95	0,986	0,928
Vegetação Natural	0,939	0,957	0,998	0,948
Florestas plantadas	0,957	0,931	0,99	0,99
Solo exposto	0,895	0,868	0,981	0,881
Corpos d'água	1	1	1	1

Fonte: o autor.

Por meio da análise da Tabela 9 verifica-se que o modelo de árvore de decisão apresenta uma taxa de acerto superior a 90% para um nível de pré-poda menor ou igual a 350, isto é, 350 *pixels* por folha da árvore ou regra de decisão gerada. Ainda analisando-se a Tabela 9 é possível verificar que um ponto de corte ótimo poderia

ser aquele resultante da aplicação de um nível de pré-poda igual ou menor a 100 objetos por folha, visto que a acurácia é mantida acima de 91% para estes níveis de pré-poda. A partir desse ponto de corte, a acurácia tende a diminuir mais rapidamente.

Tabela 9 -Taxa de acerto, a estatística Kappa e o número de regras para diferentes níveis de pré-poda.

Número mínimo de objetos por folha	Taxa de acerto (%)	Estatística Kappa	Regras
5	96,47	0,95	373
10	96,19	0,95	274
20	95,71	0,94	167
30	95,21	0,94	123
40	94,66	0,93	94
50	94,52	0,93	84
60	94,30	0,92	74
70	94,31	0,92	65
80	94,21	0,92	61
90	94,07	0,92	60
100	93,81	0,92	57
200	91,71	0,89	40
300	91,30	0,88	30
400	89,92	0,86	23
500	89,33	0,86	19
600	89,26	0,86	18
700	88,88	0,85	17
2500	81,59	0,76	6

Fonte: O autor.

Outra informação relevante da Tabela 9 é que ocorre uma diminuição expressiva do número de regras para níveis de pré-poda superiores 40 objetos por folha. Para este número de objetos por folha observa-se que o número de regras cai para um patamar de 70%, de modo que a acurácia ainda se aproxima dos 95%.

4 CONCLUSÕES

As técnicas de mineração de dados utilizadas neste trabalho se mostram eficientes na resolução de problemas de identificação de florestas plantadas em imagens de satélite, tanto pelo desempenho da classificação, como pela redução da quantidade de informação necessária para a resolução deste tipo de problema.

Os resultados obtidos por meio da aprendizagem por árvore de decisão demonstraram que a utilização de um grande número de atributos obtidos a partir das bandas do sensor escolhido foi vantajosa para o Algoritmo Classificador, sendo que os métodos de seleção de atributos identificaram subconjuntos reduzidos de atributos.

Embora os atributos de textura e índices de vegetação não tenham trazido ganho no resultado da classificação realizada os modelos de árvore de decisão binária, por

meio do algoritmo J48, alcançaram taxas de acerto superiores a 90% na identificação das espécies plantadas em meio a outros alvos.

A estatística *Kappa*, principal medida de qualidade do modelo foi 0,92, em um índice que varia entre 0 e 1, o que constitui uma forte evidência do potencial dos modelos de árvore de decisão no processo de classificação de florestas plantadas em imagens de satélite.

5 REFERÊNCIAS

ADAPEC -TO, **Agência de Defesa Agropecuária do Tocantins**. Lista de imóveis com reflorestamentos por município no Estado do Tocantins/TO. Palmas, 2013.

BATISTA, M. H.; HAERTEL, V. **Classificação Hierárquica Orientada a Objetos em Imagens de Alta Resolução Espacial Empregando Atributos Espaciais e Espectrais**. 2006. 216p. Dissertação (Mestrado em Sensoriamento Remoto) - Universidade Federal do Paraná (UFPR). Porto Alegre-PR.

CELINSKI, Tatiana Montes. **Classificação de cobertura do solo utilizando árvores de decisão e sensoriamento remoto**. 2008. 126p. Tese (Doutorado em Agronomia (Energia na Agricultura)) - UNIVERSIDADE ESTADUAL PAULISTA.

- CHANDER, G., & MARKHAM, B. (2003). **Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges**. *Geoscience and Remote Sensing, IEEE Transactions*, 41(11), 2674-2677.
- DAMASO, M. C. T.; HELM, C. V.; SILVA, P. R.; CONTE, R. A.; GAMBETTA, R. **Pesquisa para turbinar biomassa florestal**. *Revista Ambiente Energia*, ano 1, n. 2, p. 19-22, out. /dez. 2013. Disponível em: <<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/93518/1/Monica-Revista.pdf>>. Acesso em 15 abr. 2014
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases**, *AI magazine*, v.17, n. 3, p. 37-54, 1996.
- FIORESE, G.; GUARISO, G. **A GIS-based approach to evaluate biomass potential from energy crops at regional scale**. *Environmental Modelling & Software*, v. 25, n. 6, p.702-711,2010. ISSN 1364-8152. Disponível em: <<http://www.sciencedirect.com/science/article/B6VHC-4XW6SY81/2/e7a8d3c90f9f822406da324d7ec9089d>>. Acesso em:16 nov. 2013.
- HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques**. San Francisco: Morgan Kaufmann Publishers, 770p, 2006.
- HARALICK, R.M; SHUNMUGAN, K; DINSTEIN, I. **Texture Feature for Image Classification**. *IEEE Transactions on Systems, Man and Cybernetics*, v. 3, n. 6, 610-621, Nov. 1973.
- LEITE, N. B. **Novas fronteiras: desafios e oportunidades**. *Revista Opiniões*. Disponível em: <<http://revistaonline.revistaopinioes.com.br/revistas/flo/59/#page/5>>. Acesso em: 15 nov. 2012.
- LENZ, A. R. **Utilizando técnicas de aprendizado de máquina para apoiar o teste de regressão**. 2009. 150p. Dissertação (Mestrado em Informática) - Universidade Federal do Paraná (UFPR).
- NONATO, R. T. **Aplicação de mineração de dados na identificação de áreas cultivadas com cana-de-açúcar em imagens de sensoriamento remoto no estado de São Paulo**. 2010. 128 p. Dissertação (Mestrado em Engenharia Agrícola) - Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas.
- PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento remoto no estudo da vegetação**. São José dos Campos. Editora Parentese, 2007.
- QUINLAN. J. R. **C4.5: programs for machine learning**. Morgan Kaufmann PublishersInc., San Francisco, CA, USA, 1993.
- RAMACHANDRA, T. V. RIEP: Regional integrated energy plan. **Renewable and Sustainable Energy Reviews**, v.13, n.2, p. 285-317, 2009. ISSN 13640321. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1364032107001414>>. Acesso em 02 mai. 2014.
- SEMADES-TO. Secretaria de Meio Ambiente e Desenvolvimento Sustentável do Estado do Tocantins: **Reformulação da Política Estadual de Florestas e Elaboração do Plano Estadual de Florestas do Tocantins**, Palmas, 2013.
- TANAJURA, E. L. X., Antunes, M. A. H., & Uberti, M. S. (2005). **Avaliação de índices de vegetação para a discriminação de alvos agrícolas em imagens de satélites**. *Simpósio Brasileiro de Sensoriamento Remoto, 12 (SBSR)*, 301-308.
- VASCONCELOS, Simone. **Matriz de Co-ocorrência**. 2010. Disponível em: <<http://www.ic.uff.br/~aconci/co-ocorrencia.pdf>> Acesso em 15 abr. de 2014.
- WITTEN, I. H.; FRANK, E.; HALL, M. A. **Data mining: practical machine learning tools and techniques**. 3rd ed. San Francisco: Morgan Kaufmann, 2011. 629 p.