CARLOS TAVARES NONATO
YOLANDA VIEIRA DE ABREU

LIFE EDITORA

DESENVOLVIMENTO DE MODELOS DE CLASSIFICAÇÃO DE FLORESTAS PLANTADAS PARA PLANEJAMENTO E TOMADA DE DECISÃO: MINERAÇÃO DE DADOS E IMAGENS DE SATÉLITE

NÚCLEO EM INTERUNIDADES EM DESENVOLVIMENTO ECONÔMICO, SOCIAL E ENERGÉTICO.UFT-TO-BR

MESTRADO EM AGROENERGIA DA UFT - TO - BRASIL

Digitized by Google

Conselho Editorial Life Editora

Prof. Dr. Amilcar Araujo Pereira
UFRJ/ Faculdade de Educação

Prof. Dr. Edgar César Nolasco
UFMS/ Campo Grande-MS

Prof. Dr. Gilberto José de Arruda
UEMS/ Unidade de Dourados

Prof. Dr. Matheus Wemerson G. Pereira
UFMS/Campo Grande-MS

Prof. Dr. Giovani José da Silva
UFMS/Campus de Nova Andradina

Profª. Dra. Helena H. Nagamine Brandão
Universidade de São Paulo - USP-SP

Profª. Dra. Joana Aparecida Fernandes Silva
UFG/Goiás

Prof. Dr. João Wanderley Geraldi
Universidade do Porto, Portugal/ INEDD/Universidade Siegen/Alemanha e Unicamp

Profª. Dra. Léia Teixeira Lacerda
UEMS/ Unidade de Campo Grande

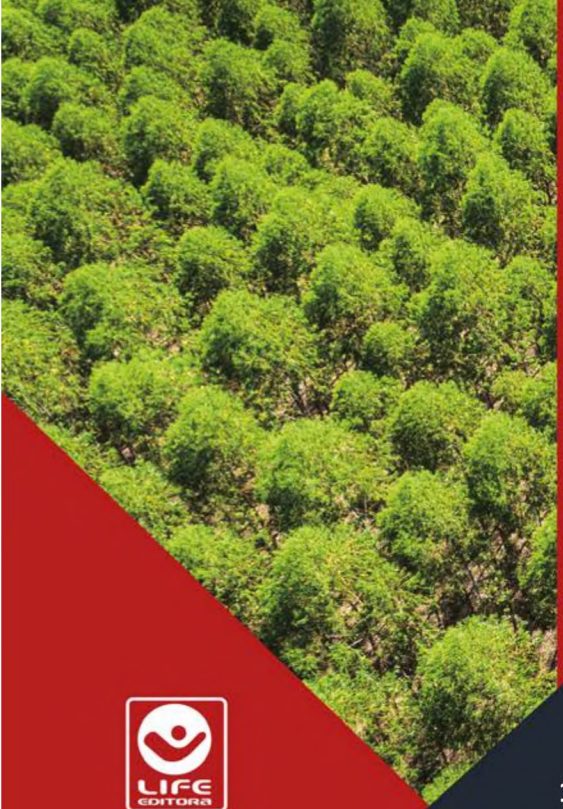
Profª. Dra. Maria Cecília Christiano Cortez de Souza
Faculdade de Educação da Universidade de São Paulo

Profª. Dra. Maria Leda Pinto
UEMS/ Unidade de Campo Grande

Prof. Dr. Marlon Leal Rodrigues
UEMS/ Unidade de Campo Grande

O papel utilizado em nossos livros são biodegradáveis e renováveis. Provém de florestas plantadas que dão emprego a milhares de brasileiros e combatem o efeito estufa, pois absorvem gás carbônico durante o seu crescimento! A tinta utilizada na impressão das páginas é à base de soja, cujo componente é renovável e atóxico que não degrada o meio ambiente.

Digitized by Google



CARLOS TAVARES NONATO
YOLANDA VIEIRA DE ABREU

1ª Edição - Campo Grande - MS - Brasil - 2020

DESENVOLVIMENTO DE MODELOS DE CLASSIFICAÇÃO DE FLORESTAS PLANTADAS PARA PLANEJAMENTO E TOMADA DE DECISÃO: MINERAÇÃO DE DADOS E IMAGENS DE SATÉLITE

NÚCLEO EM INTERUNIDADES EM DESENVOLVIMENTO ECONÔMICO, SOCIAL E ENERGÉTICO.UFT-TO-BR

MESTRADO EM AGROENERGIA DA UFT - TO - BRASIL

Digitized by Google

Copyright © by **Carlos Tavares Nonato**
Yolanda Vieira de Abreu

Direitos Autorais reservados de acordo com a Lei 9.610/98

Coordenação Editorial
Valter Jeronymo

Revisão
Os próprios autores

Assistente de Coordenação
Raquel de Souza

Formatação
Life Digital

Projeto Gráfico
Life Editora

Associado

Life Editora
Rua Américo Vespúcio, 255 - Santo Antonio
CEP: 79.100-470 - Campo Grande - MS
Fones: (11) 3508 1941 - Cel.: (67) 99297-4890
contato@lifeeditora.com.br • **www.lifeeditora.com.br**

lifeeditora
lifeeditora
lifeeditora
lifeeditora
lifeeditora
lifeeditora

Dados Internacionais de Catalogação na Publicação (CIP)

Nonato, Carlos Tavares
Abreu, Yolanda Vieira de

Desenvolvimento de modelos de classificação de florestas plantadas para planejamento e tomada de decisão: Mineração de Dados e Imagens de Satélite / Carlos Tavares Nonato e Yolanda Vieira de Abreu– Campo Grande, MS: Life Editora, 2020.

124p. : il. : 23 cm

ISBN 978-65-5887-013-5

1. Florestas plantadas 2. Mineração de dados I. Título

CDD - 360

Proibida a reprodução total ou parcial, sejam quais forem os meios
ou sistemas, sem prévia autorização dos autores.

Digitized by Google

Dedicamos ao Miguel Martins Dias Nonato,
representando as futuras gerações.

Digitized by Google

Agradecimentos

A Deus pela presença constante em minha vida.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro no início deste trabalho.

À ADAPEC pela disponibilização de dados referentes aos imóveis com reflorestamentos no estado do Tocantins.

A todas as pessoas que direta ou indiretamente me auxiliaram na execução e aprimoramento deste trabalho.

Digitized by Google

Lista de Símbolos e Abreviações

ADAPEC: Agência de Defesa Agropecuária do Estado do Tocantins
CBERS: *China-Brazil Earth Resources Satellite*
(Satélite Sino-Brasileiro de Recursos Terrestres)
CCD: *Charge-coupled Device* (Câmera Imageadora de Alta Resolução)
CTVI: *Corrected Transformed Vegetation Index*
(Índice de Vegetação Transformado Corrigido)
DN: *Digital Number* (Número Digital)
EMBRAPA: Empresa Brasileira de Pesquisa Agropecuária
FAO: *Food and Agriculture Organization of the United Nations*
(Organização das Nações Unidas para a Agricultura e a Alimentação)
GEE: Gás Efeito Estufa
GEMI: *Global Environment Monitoring Index*
(Índice Global de Monitoramento Ambiental)
IAF: Índice de área foliar
IBGE: Instituto Brasileiro de Geografia e Estatística
iLP: Integração Lavoura Pecuária
iLPF: Integração Lavoura Pecuária Floresta
INPE: Instituto Nacional de Pesquisas Espaciais
LANDSAT: *Land Remote Sensing Satellite System*
(Satélite de Sensoriamento Remoto Terrestre)
MAXVER: Máxima Verossimilhança
MLME: Modelo linear de mistura espectral
ONU: Organização das Nações Unidas
REM: Radiação eletromagnética
SAFs: Sistemas Agroflorestais
SIG: Sistema de Informações Geográficas
SR: *Simple Ratio* (Razão Simples)
SPD: Sistema de Plantio Direto
TM: *Thematic Mapper*
TVI: *Transformed Vegetation Index* (Índice de Vegetação Transformado)

Digitized by Google

Nota dos Autores

As florestas plantadas tem atraído grande interesse pela possibilidade de utilização em aplicações bioenergéticas frente à tendência mundial de priorizar fontes de energia que proporcionem maior sustentabilidade ambiental, mais qualidade e segurança. No Brasil, os deslocamentos na geografia da cadeia produtiva agroflorestal atual em direção às regiões de fronteira agrícola (Centro-Oeste e Norte) vem criando desafios de adequação dos conhecimentos técnico-científicos já consolidados em outras regiões. Nesse contexto, o objetivo desta dissertação é avaliar a acurácia da classificação e identificação de áreas cultivadas com florestas plantadas para fins energéticos, em imagens orbitais do sensor Landsat 5 TM. Por meio de técnicas estatísticas de mineração de dados, o presente trabalho também avaliou a utilização de um amplo conjunto de atributos para identificar melhorias nos resultados da classificação. A pesquisa se concentrou em amostras de áreas plantadas no estado do Tocantins, região norte do Brasil. 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, como pela redução da quantidade de informação necessária para a resolução deste tipo de problema. Assim, as técnicas empregadas nesse estudo possibilitam o desenvolvimento de modelos de classificação robustos no auxílio ao planejamento e à tomada de decisão sobre a plantação de florestas no território brasileiro.

Estudo defendido como dissertação no Mestrado em Agroenergia da Universidade Federal do Tocantins em 2014. Os autores fizeram pequenas modificações ao transformar em livro

Digitized by Google

Sumário

CAPÍTULO 1: INTRODUÇÃO.....11

CAPÍTILO 2: REFERENCIAL TEÓRICO.....15
2.1 Sensoriamento Remoto.....15
2.1.1 Interação da radiação eletromagnética com os alvos.....17
2.1.2 O comportamento espectral da vegetação.....19
2.1.3 Índices de vegetação.....21
2.1.4 Resoluções das imagens.....23
2.1.5 O Satélite Landsat 5 e o sensor Thematic Maper (TM).....24
2.2 Descoberta de conhecimento em bancos de dados.....32
2.3 Fases da descoberta de conhecimento em bancos de dados (KDD).....35
2.4 Mineração de dados.....39
2.4.1 Principais tarefas de mineração de dados.....40
2.5 Árvores de decisão.....45
2.5.1 Regras de classificação.....46
2.5.2 Avaliação da Classificação.....47
2.5.3 Estatística Kappa de Cohem e a Análise de Concordância.....50

CAPÍTULO 3: PANORAMA DAS FLORESTAS PLANTADAS NO MUNDO, NO BRASIL E NO TOCANTINS.....52
3.1 Panorama do setor florestal no mundo.....52
3.2 Panorama do setor florestal no Brasil.....54
3.3 Florestas plantadas e energia.....58
3.1.1 Florestas energéticas no Tocantins.....62
3.4 Plano ABC.....67

CAPÍTULO 4: PROCEDIMENTOS METODOLÓGICOS.....75
4.1 Métodos.....75
4.1.1 Dados de sensoriamento remoto utilizados no trabalho.....76
4.1.2 Área de estudo.....77
4.2 Seleção das áreas de estudo (Fase 1).....77
4.2.1 Estimativa das áreas com florestas plantadas e o plano amostral.....78
4.2.2 Lista final de municípios.....86
4.3 Processamento das imagens (Fase 2).....87
4.3.1 Correção radiométrica.....87
4.3.2 Conversão de Números Digitais (NDs) para valores de reflectância.....87

Digitized by Google

4.4 Preparação dos dados (Fase 3).....89
4.4.1 Extração dos dados relativos às regiões de interesse.....89
4.4.2 Transformação dos dados da imagem para o formato tabular.....89
4.4.3 Consolidação dos dados em arquivos individuais.....89
4.5 Mineração de dados (Fase 4).....90
4.5.1 Definição dos atributos e entendimento dos dados.....91
4.5.2 Seleção de atributos.....93
4.5.3 Aplicação do modelo de classificação.....93
4.5.4 Avaliação dos resultados da classificação.....95

CAPÍTULO 5: RESULTADOS E DISCUSSÕES.....96
5.1 Seleção inicial de atributos.....96
5.2 Avaliação do modelo de classificação de maior acurácia.....98
5.3 Avaliação de diferentes níveis de pré-poda para o modelo de melhor acurácia...100
5.4 Considerações sobre os resultados obtidos.....104

CONCLUSÕES E RECOMENDAÇÕES.....105

REFERÊNCIAS BIBLIOGRÁFICAS.....107

APÊNDICE I: Rotina de classificação do modelo de melhor acurácia.....116

ANEXO I: Lista de imóveis com reflorestamentos por município no Estado do Tocantins, coordenadas geográficas e a área utilizada em cada propriedade...119

Digitized by Google

CAPÍTULO 1:

INTRODUÇÃO

A energia derivada da biomassa cultivada, também chamada de agroenergia, tem ganhado grande destaque como fonte renovável devido à tendência mundial de priorizar tecnologias que proporcionem sustentabilidade ambiental, melhor qualidade e maior segurança no fornecimento de energia. Esta tendência é acentuada pela perspectiva de aumento da demanda de combustíveis, a instabilidade do preço do petróleo e pelos temores de que as mudanças climáticas e ambientais sejam irreversíveis (MIURA, 2012). No entanto, ainda não há saídas tecnológicas economicamente viáveis para substituir toda a cadeia de produtos derivados do petróleo. Diante dessa realidade, a integração fóssil/renovável é uma solução que se apresenta cada vez mais próxima da nossa realidade e o Brasil pode ocupar papel de destaque no cenário mundial por possuir grandes áreas para a plantação e áreas degradadas que podem ser recuperadas, além de condições climáticas e ambientais favoráveis (DAMASO *et al.*, 2013).

Os investimentos em florestas plantadas tem concentrado grande interesse pela possibilidade de 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).

Segundo Leite (2012), pode-se notar uma mudança na geografia da cadeia produtiva agroflorestal atual. 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.

Digitized by Google

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. Há necessidade de contemplar a agroenergia em estudos territoriais e no planejamento energético do país, pois segundo Ramachandra (2009), quando as necessidades energéticas estão integradas às ações sociais, políticas, econômicas e às limitações ambientais, tanto em nível local quanto global, é possível alcançar o desenvolvimento sustentável para uma região.

Nesse sentido, as geotecnologias têm sido amplamente utilizadas para responder as demandas relativas ao planejamento bioenergético (FIORESE; GUARISO, 2010; GRAHAM *et al.*, 2000), pois permitem que múltiplas e diferentes variáveis possam ser correlacionadas espacialmente e temporalmente. A visualização de diferentes variáveis que influenciam na dinâmica da produção bioenergética é um instrumento fundamental à tomada de decisão por parte de gestores do setor bioenergético.

Como definição, as geotecnologias podem ser entendidas como o conjunto de tecnologias para a coleta, processamento, análise e disponibilização de informação com referência geográfica. São originadas de estudos e pesquisas na área de sensoriamento remoto (NASCIMENTO e ABREU, 2012).

Segundo Hamachandra (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. Essas aplicações abrangem diversas etapas como, por exemplo, o processamento de imagens e a classificação dos fenômenos encontrados na área abrangida pela imagem. Na etapa de classificação busca-se rotular cada *pixel* ou elemento da imagem segundo a ocupação existente no solo, obtendo-se um mapa de classes temáticas.

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 min-

Digitized by Google

eração 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 (CELINK, 2008).

Este estudou utilizou-se dessa técnica de mineração de dados e a árvore de decisão foi avaliada na classificação de pixels puros de regiões com florestas plantadas em meio a alvos diversos, como áreas urbanas, solo exposto, corpos d'água, e outros tipos de vegetação.

O estudo de caso foi realizado com dados pertinentes ao norte do estado do Tocantins, que se destaca por concentrar a maior parte dos plantios de florestas destinados a produção de energia.

A pergunta que pauta a elaboração deste trabalho é: Qual a aderência das técnicas de mineração de dados na resolução de problemas de classificação de áreas cultivadas com florestas plantadas em imagens de sensoriamento remoto? Para responder a essa questão, dois aspectos foram investigados: 1. O potencial da técnica de árvore de decisão na classificação de áreas cultivadas com florestas plantadas em imagens de sensoriamento remoto orbital; 2. a identificação do conjunto de atributos mais significativo para o cenário estudado.

O objetivo principal deste estudo 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 via técnicas de mineração de dados. A região norte do estado do Tocantins foi selecionada como estudo de caso para responder à questão central do trabalho.

Os caminhos escolhidos para chegar à solução do problema e do objetivo principal foram os seguintes:

- Discutir a importância das florestas plantadas para fins energéticos no contexto socioeconômico do território brasileiro e, especificamente no estado do Tocantins. Consequentemente, demonstrar a importância de pesquisas relativas à identificação e classificação de dados que nor-

Digitized by Google

teiam o planejamento desse tipo de cultivo.

- Analisar e descrever os dados de área plantada com florestas no Tocantins, obtidos junto a Agência de Defesa Agropecuária do Estado do Tocantins (ADAPEC).
- Analisar dentro da técnica de mineração de dados o potencial do algoritmo de árvore de decisão J48 na classificação de áreas de florestas plantadas em imagens de sensoriamento remoto orbital, especificamente do satélite Landsat-5, sensor TM.
- Desenvolver considerações sobre a utilização do método de árvore de decisão na classificação de áreas com florestas plantadas em imagens de sensoriamento remoto.

Esse estudo foi dividido em seis capítulos, sendo este o primeiro. O segundo versa sobre a fundamentação teórica sobre sensoriamento remoto, aplicação dos índices de vegetação e características dos sensores remotos. O terceiro capítulo aborda o processo de descoberta de conhecimento em bancos de dados ou KDD (*Knowlegde Discovery in Databases)*, suas respectivas fases e o conceito de mineração de dados com ênfase para os algoritmos de árvore de decisão. O quarto capítulo descreve o panorama das florestas plantadas no Mundo, no Brasil e no Tocantins, a cadeia de florestas plantadas para fins energéticos e o plano ABC. O quinto descreve a metodologia, os materiais utilizados e a aplicação dos modelos de classificação. O sexto contém a os resultados e a conclusão.

Digitized by Google

CAPÍTILO 2:

REFERENCIAL TEÓRICO

2.1 Sensoriamento Remoto

A definição mais conhecida de sensoriamento remoto o descreve como uma técnica de obtenção de imagens de objetos da superfície terrestre sem que haja contato físico de qualquer espécie entre o sensor e o objeto (FLORENZANO, 2002 *apud* NASCIMENTO, 2011; MENESES e ALMEIDA, 2012).

Segundo Menezes e Almeida (2012), essa tecnologia que utiliza instrumentos capazes de obter imagens da superfície terrestre a distâncias remotas pode ser definida de maneira mais crítica como uma ciência que visa o desenvolvimento da obtenção de imagens da superfície terrestre por meio da detecção e medição quantitativa das respostas das interações da radiação eletromagnética com os materiais terrestres.

Meneses e Almeida (2012) destacam que essa definição de sensoriamento leva em consideração premissas como a ausência de matéria no espaço entre o objeto e o sensor, a capacidade de transportar a informação do objeto pelo espaço vazio e a comunicação entre objeto e sensor por meio da radiação eletromagnética, sendo essa a única forma de energia capaz de se transportar pelo espaço. Desse modo o entendimento do processo de aquisição e de análise de imagens perpassa inicialmente pelo conhecimento do que é a radiação eletromagnética - REM.

No sensoriamento remoto a energia ou onda eletromagnética natural da superfície do planeta Terra é medida por sensores localizados a bordo de satélites artificiais de observação da terra ou de aviões. Essa energia captada pelos sensores pode ser a refletida, que é a energia proveniente do sol e tem interação com a superfí-

Digitized by Google

cie terrestre, ou pode ser a energia emitida pela própria superfície terrestre (MENESES e ALMEIDA, 2012).

Os sensores a bordo dos satélites são sensíveis a um número variado de comprimentos de onda da radiação eletromagnética. Usualmente, os comprimentos de onda são divididos em bandas diferentes que cobrem faixas específicas do espectro eletromagnético.

Os diferentes objetos que fazem parte da cobertura da superfície terrestre como a água, a vegetação e áreas urbanas, absorvem e transmitem quantidades de energia eletromagnética em diferentes proporções que variam com o comprimento de onda, de modo que a energia refletida varia nas diferentes bandas do espectro eletromagnético, isto é, as diferentes coberturas de solo produzem diferentes respostas espectrais (MENESES e ALMEIDA, 2012).

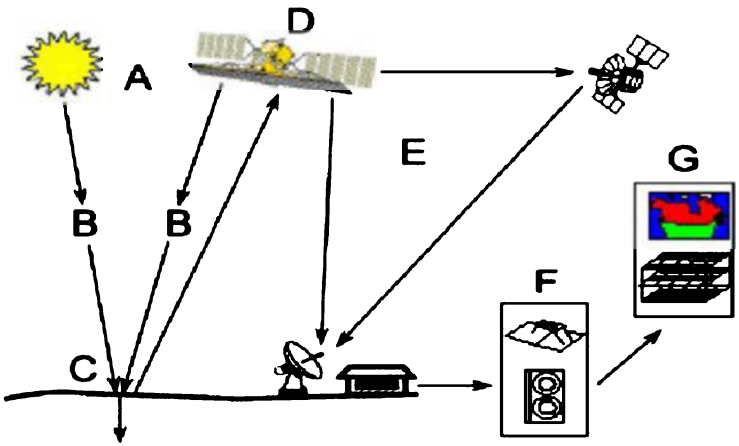
Essas diferentes respostas fornecidas pelos objetos caracterizam o comportamento espectral dos alvos e são as principais fontes de dados em estudos voltados para o monitoramento de biomassa (BATISTA e AFFONSO, 2002).

Os sistemas de imageamento e coleta de dados por satélite, comumente empregados no monitoramento da agricultura podem ser observados na Figura1, e envolvem diferentes elementos e procedimentos, sendo os mais relevantes:

- Fontes de energia (A)
- Radiação e atmosfera (B)
- Interação da radiação com o alvo (C)
- Gravação da energia pelo sensor (D)
- Transmissão, recepção e processamento (E)
- Interpretação e análise (F)
- Aplicação (G)

Digitized by Google

Figura 1 – Processo de imageamento e coleta de dados em sensoriamento remoto



Fonte: Geoeasy, 2011

2.1.1 Interação da radiação eletromagnética com os alvos

Durante sua passagem através da atmosfera, a energia eletromagnética que vem do Sol ou a transmitida pela Terra interage com elementos gasosos e partículas suspensas na atmosfera que interferem na intensidade da radiação, na distribuição espectral e na direção dos raios incidentes. Os dois efeitos dessa interação da radiação com a atmosfera são denominados absorção e espalhamento (SLATER, 1980; MENEZES e ALMEIDA, 2012).

Segundo Meneses e Almeida (2012) a absorção é o efeito mais prejudicial ao sensoriamento remoto, tal que moléculas e aerossóis na atmosfera absorvem energia em vários comprimentos de onda e como consequência a radiação solar pode ter sua intensidade afetada ou até ser impedida de atingir a superfície terrestre. Já no espalhamento ocorre uma mudança aleatória da direção de propagação da radiação solar, conforme o comprimento de onda, a quantidade de partículas ou gases e a distância percorrida pela REM e tem como consequência a diminuição da eficiência do sensoriamento remoto na identificação dos alvos.

Digitized by Google

Meneses e Almeida (2012) destacam também que nas regiões onde não ocorre absorção total ou intensa da REM, também conhecidas como janelas atmosféricas, podem ocorrer interações da radiação eletromagnética com os alvos terrestres por meio de processos de absorção, onde a radiação é absorvida pelo alvo, transmissão, onde a radiação atravessa o alvo e/ou reflexão onde a radiação é refletida e redirecionada.

A proporção na qual ocorrem as interações com a superfície depende do comprimento de onda e da composição bio-físico-química dos alvos, sendo a reflexão considerada a interação mais importante para o sensoriamento remoto, de modo que a partir da energia refletida ou emitida pelos objetos, podem-se obter padrões de respostas espectrais que permitem distinguir os alvos terrestres (SLATER, 1980).

Em estudos que utilizam o sensoriamento remoto, para que se possa explicar o que se observa nas imagens, é extremamente importante a compreensão da existência da radiação eletromagnética em forma de onda e na forma de energia.

O espectro eletromagnético foi arbitrariamente dividido em intervalos de comprimento de onda em função do uso que o homem encontrou para suas aplicações. Para os usuários do sensoriamento remoto é essencial saber os valores dos comprimentos de onda desses intervalos, os nomes que recebem e quais os comprimentos de onda são possíveis de ser detectados por cada tipo de sensor (MENEZES e ALMEIDA, 2012).

A Tabela 1 apresenta os intervalos de comprimento de onda do espectro eletromagnético, com destaque para a faixa do espectro visível e infravermelho.

Como pode ser observado na Tabela 1 o espectro eletromagnético abrange desde curtos comprimentos de onda, como os raios cósmicos e os raios gama, de alta frequência, até longos comprimentos de ondas de rádio e TV, de baixa frequência

As faixas espectrais destacadas em negrito são as mais utilizadas no sensoriamento remoto. Na região do visível, o olho humano

Digitized by Google

enxerga a energia (luz) eletromagnética, sendo capaz de distinguir as cores do violeta ao vermelho. A região do infravermelho é subdividida em quatro regiões: infravermelho próximo (0,76 -1,2 µm), de ondas curtas (1,2 - 3,0 µm), médio (3,0 - 5,0 µm) e distante ou termal (5,0 µm – 1,0mm). (FLORENZANO; 2002, MENESES; 2012).

Tabela 1 - Divisão do espectro eletromagnético

Intervalo espectral	Comprimento de onda
Raios cósmicos	0,01 Aº
Raios gama	0,01-0,1 Aº
Raios X	0,1-10 Aº
Ultravioleta	100nm -0,38µm
Visível	0,38-0,76 µm
Infravermelho próximo	0,76-1,2 µm
Infravermelho de ondas curtas	1,2-3,0 µm
Infravermelho médio	3,0-5,0 µm
Infravermelho termal	5,0-1 µm
Micro-ondas	1 nm -100 cm
Rádio	1m – 10 km
Áudio	10 – 100 km
Corrente alternada	> 100 km

*Obs: Aº = 10^{-10} m; nm = 10^{-9} m; µm = 10^{-6}m
Fonte: Meneses e Almeida (2012)

2.1.2 O comportamento espectral da vegetação

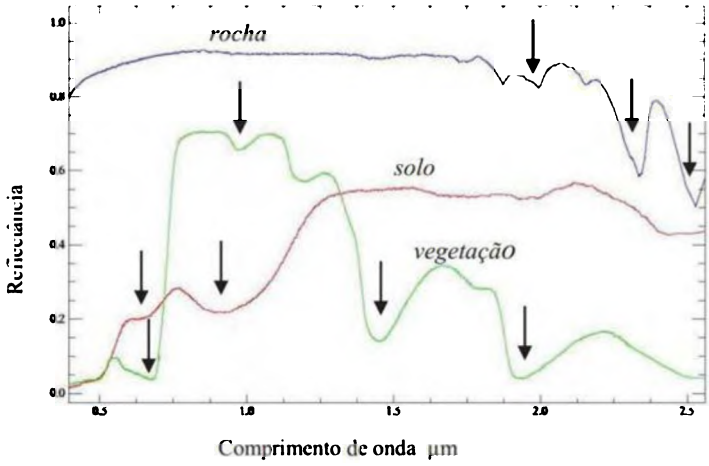
O conhecimento da interação entre a vegetação e o meio que a cerca tem grande importância para o entendimento dos fenômenos naturais, pois a vegetação exerce um papel importante nos ciclos hidrológico e biogeoquímico da superfície terrestre (PEREIRA *et al.*,1996).

Estudos sobre a relação entre as características da vegetação e dados coletados remotamente tem sido frequentemente aplicados no setor agrícola e florestal (PONZONI e SHIMABUKURO, 2007).

Digitized by Google

A Figura 2 mostra a relação entre a porcentagem de reflectância e o comprimento de onda. A observação da Figura 2 permite identificar em quais comprimentos de onda ocorrem as feições de absorção dos materiais rocha, solo e vegetação. No caso dos dosséis, que são os estratos superiores da floresta ou de culturas agrícolas, constituídos por elementos como galhos, folhas, frutos e flores, as propriedades de recfletância dependem principalmente das características da folha. Os principais mecanismos biológicos que influenciam o comportamento espectral, em geral, são: absorção pelos pigmentos (baixa ~10% na faixa do visível), estrutura celular interna da folha (elevada ~50% no infravermelho próximo) e concentração de água nos tecidos. Estes mecanismos biológicos respondem de forma distinta de acordo com a disponibilidade de água no ambiente, idade e saúde da planta. Além disso, a reflectância dos dosséis pode ser afetada por fatores físicos como o ângulo de incidência da radiação, ângulo de visa e arquitetura da planta (PEREIRA; 1996).

Figura 1 – Exemplo de comportamento espectral de vegetação, rocha e solo na superfície terrestre



Fonte: Meneses e Almeida (2012)

Segundo Ponzoni e Shimabukuro (2002), em seu trabalho realizado sobre o sensoriamento remoto no estudo da vegetação,

Digitized by Google

as imagens referentes aos dosséis na região do visível apresentam tonalidade escura devido à baixa reflectância da REM, em função dos pigmentos fotossintetizantes. Já em imagens da região do infravermelho próximo, os dosséis apresentam a tonalidade clara e nas imagens do infravermelho médio espera-se tons de cinza intermediários.

Desse modo cada dossel em particular, possui características únicas e desenvolve-se em diferentes tipos de solo e sob diferentes condições ambientais e edafoclimáticas. Assim como acontece em outros objetos de estudo do sensoriamento remoto, podem ocorrer desvios nos quais diferentes parâmetros podem assumir valores iguais de radiância, o que acarretará em uma mesma aparência nas imagens, mesmo em se tratando de diferentes coberturas vegetais. Neste caso cabe ao interprete conhecer bem estas limitações e buscar extrair dados confiáveis das imagens de sensoriamente remoto.

2.1.3 Índices de vegetação

Para minimizar a variabilidade causada por fatores externos como solo, atmosfera e geometria de aquisição de dados, a reflectância espectral tem sido transformada e combinada em vários indices de vegetação. Esta abordagem para a definição de parâmetros de vegetação é referida na literatura como abordagem empírica (HALL *et al*, 1995).

Os índices de vegetação mais comuns utilizam a informação contida na reflectância dos dosséis referente às regiões do vermelho e infravermelho próximo, as quais são resultados de combinações de duas ou mais bandas espectrais através da soma, diferença, razão entre as bandas ou qualquer outra combinação (WIEGAND *et al.*, 1974 citado por SHIMABUKURO *et al.*, 2009).

Segundo Chen *et al.* (1986), o emprego dos indices de vegetação para caracterizar e quantificar determinado parâmetro biofisico de culturas agricolas, tem duas grandes vantagens: a) permitir

Digitized by Google

reduzir a dimensão das informações multiespectrais através de um simples número, além de minimizar o impacto das condições de iluminação e visada; b) fornece um número altamente correlacionado aos parâmetros agronômicos.

No entanto, os vários indices de vegetação podem ser diferentemente afetados pelas características de iluminação e visada e pela arquitetura do dossel e pelo substrato abaixo do dossel, justificando assim um estudo para avaliar o tipo de indice de vegetação mais adequado para cada aplicação. Para floresta de coníferas como o pinus, MacDonald *et al.*, (1998) *apud* Tanajura *et al.*, (2005) encontraram que os indices de vegetação são significantemente afetados por ângulo zenital solar, reflectância do substrato, estrutura do dossel e IAF.

Indices de vegetação podem ser divididos em dois grupos (BARET e GUYOT, 1991 *apud* TANAJURA *et al.*, 2005): os índices baeados em inclinação, como o NDVI e os baseados na distância como o PVI.

O índice de vegetação da diferença normalizada (NDVI) e as imagens-fração derivadas do modelo linear de mistura espectral (MLME), dentre várias técnicas de transformações radiométricas disponíveis, têm sido amplamente utilizados para gerar informações da cobertura florestal (SHIMABUKURO *et al.*, 2009). O índice de vegetação melhorado (EVI) foi desenvolvido para ter maior sensibilidade em regiões com alta produção de biomassa com redução da influência atmosférica e do solo. No Brasil, o NDVI, EVI e os demais índices têm sido amplamente aplicados por pesquisadores e silvicultores interessados em conhecer melhor os processos ecofisilógicos que controlam suas florestas (ALVARES *et al.*, 2013). A Tabela 2 apresenta os principais indices de vegetação uilizados no monitoramento de biomassa e agricultura e suas respectivas relações funcionais entre as bandas espectrais.

Digitized by Google

Tabela 2- Índices de vegetação, fórmulas e referências

Índice	Fórmula	Referência
NDVI	(B4-B3)/ (B4+B3)	Rouse *et al.*, (1974)
PVI	$((B3 - B3_{solo})^2 + (B4 - B4_{solo})^2)^{1/2}$	Richardson e Wiegand, (1977)
SAVI	$((1 + L) * (B4_{solo} - B3_{solo})) / (B4_{solo} + B3_{solo} + L)$	Huete, (1988)
RVI	B4/B3	Jordan, (1969)
EVI	G*(B3-B4)/(L+B4+ (C1*B3) – (C2*B1))	Justice *et al.*,(1998)

Fonte: PONZONI e SHIMABUKURO (2007)

Na Tabela 2, 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 píxels do solos expostos para as bandas 3 e 4 respectivamente e L é uma constante igual a 0,5.

- NDVI – Normalized Difference Vegetation Index
- PVI – Perpendicular Vegetation Index
- SAVI – Soil Adjusted Vegetation Index
- RVI – Ratio Vegetation Index

2.1.4 Resoluções das imagens

Atualmente, o sensoriamento remoto é constituído por uma grande quantidade de satélites que oferecem imagens para atender as necessidades de uma ampla gama de usuários. Para pesquisas que necessitam identificar o tamanho e a forma dos objetos há sensores que detectam áreas unitárias inferiores a 1 metro. Para pesquisas voltadas ao monitoramento da evolução e mudanças na paisagem podem recorrer aos sensores com alta taxa de revisita à área e para usuários que necessitam determinar a composição ou constituição dos minerais ou rochas, em geral, são utilizados sensores com um grande número de bandas espectrais (MENESES e ALMEIDA, 2012).

A resolução é definida por Batista e Afonso (2012) como "a

Digitized by Google

medida da habilidade que o sistema possui em distinguir objetos que estão próximos espacialmente ou respostas que são semelhantes, espectralmente".

A detecção ou identificação de um objeto nas imagens de sensoriamento remoto pode ocorrer por meio de quatro diferentes formas:

- Pela área do campo de visada do sensor (resolução espacial)
- Pelo comprimento de onda das bandas (resolução espectral)
- Pelos valores numéricos da medida de radiância do alvo (resolução radiométrica)
- Pela data em que a imagem foi tomada (resolução temporal).

2.1.5 O Satélite Landsat 5 e o sensor Thematic Maper (TM)

As imagens mais comumente aplicadas na produção de mapas de estudos da área florestal são obtidas pelos sistemas sensores orbitais TM (*Thematic Mapper*) e HVR (*High Resolution Visible)*, a bordo dos satélites LANDSAT (*Land Remote Sensing Satellite*) e SPOT (*System e Probatoire d'Observation*) respectivamente. O sensor AVHRR (*Advanced Very High Radiometer*) abordo do satélite NOAA (*National Oceanic and Atmospheric Administration Satellite*) foi desenvolvido para uso em aplicações meteorológicas e em situações excepcionais, para a produção de mapas de uso do solo, devido ao baixo custo e a boa cobertura temporal que pode ser inferior a 24 horas. Os satélites SPOT e LANDSAT possuem resolução temporal em torno de 26 e 15 dias respectivamente.

O Landsat é o programa de aquisição de imagens da superfície terrestre com maior tempo em operação. Lançado em 1972, foi o primeiro satélite não militar de observação dos recursos terrestres a fornecer imagens de alta qualidade e com uma boa resolução espacial (NASA, 2009). No início dos anos 80, entrou em funcionamento o Landsat 4, com o sensor multiespectral TM, tecnicamente mais avançado, proporcionando uma maior resolução espacial, espectral e radiométrica (de 79 a 30 m, de 4 a 7 bandas e de 6

Digitized by Google

a 8 bits). O Landsat 8, lançado em 2013 é o mais recente da série.

Como o satélite Landsat 7 está fora de operação e o Landsat 8 ainda é recente, foram utilizadas imagens do satélite Landsat 5, que dispõe de uma grande quantidade de imagens armazenadas e disponibilizas gratuitamente. Na Tabela 3 é disponibilizada uma síntese das características e principais aplicações das bandas TM do satélite Landsat 5 para melhor compreensão das variáveis utilizadas no estudo.

Tabela 3 - Características e aplicações das bandas espectrais

Banda	Intervalo espectral (µm)	Características e aplicações das bandas *TM* do sensor Landsat-5
1	0,45 – 0,52	Apresenta grande penetração em corpos de água, com elevada transparência, permitindo estudos batimétricos. Sofre absorção pela clorofila e pigmentos fotossintéticos auxiliares (carotenóides). Apresenta sensibilidade a plumas de fumaça oriundas de queimadas ou atividade industrial. Pode apresentar atenuação pela atmosfera
2	0,52 – 0,60	Apresenta grande sensibilidade à presença de sedimentos em suspensão, possibilitando sua análise em termos de quantidade e qualidade. Boa penetração em corpos de água
3	0,63 – 0,69	A vegetação verde, densa e uniforme, apresenta grande absorção, ficando escura, permitindo bom contraste entre as áreas ocupadas com vegetação (ex.: solo exposto, estradas e áreas urbanas). Apresenta bom contraste entre diferentes tipos de cobertura vegetal (ex.: campo, cerrado e floresta). Permite análise da variação litológica em regiões com pouca cobertura vegetal. Permite o mapeamento da drenagem através da visualização da mata galeria e entalhe dos cursos dos rios em regiões com pouca cobertura vegetal. É a banda mais utilizada para delimitar a mancha urbana, incluindo identificação de novos loteamentos. Permite a identificação de áreas agrícolas
4	0,76 – 0,90	Os corpos de água absorvem muita energia nesta banda e ficam escuros, permitindo o mapeamento da rede de drenagem e delineamento de corpos de água. A vegetação verde, densa e uniforme, reflete muita energia nesta banda, aparecendo bem clara nas imagens. Apresenta sensibilidade à rugosidade da copa das florestas (dossel florestal). Apresenta sensibilidade à morfologia do terreno, permitindo a obtenção de informações sobre Geomorfologia, Solos e Geologia. Serve para análise e mapeamento de feições geológicas e estruturais. Serve para separar e mapear áreas ocupadas com pinus e eucalipto. Serve para mapear áreas ocupadas com vegetação que foram queimadas. Permite a visualização de áreas ocupadas com macrófitas aquáticas (ex.: aguapé). Permite a identificação de áreas agrícolas
5	1,55 – 1,75	Apresenta sensibilidade ao teor de umidade das plantas, servindo para observar estresse na vegetação, causado por desequilíbrio hídrico. Esta banda sofre perturbações em caso de ocorrer excesso de chuva antes da obtenção da cena pelo satélite
6	10,4 – 12,5	Apresenta sensibilidade aos fenômenos relativos aos contrastes térmicos, servindo para detectar propriedades termais de rochas, solos, vegetação e água.
7	2,08 – 2,35	Apresenta sensibilidade à morfologia do terreno, permitindo obter informações sobre Geomorfologia, Solos e Geologia. Esta banda serve para identificar minerais com íons hidroxilas. Potencialmente favorável à discriminação de produtos de alteração hidrotermal.

Fonte: Divisão e geração de imagens INPE

As imagens do sensor TM do satélite Landsat 5 são consideradas imagens de média resolução espacial. Embora as análises e

Digitized by Google

os resultados abordados neste trabalho sejam sobre imagens Landsat 5 TM, a mesma metodologia poderia ser aplicada a imagens de resolução mais alta e os resultados poderiam ser comparados quanto a qualidade da classificação.

Dentre as principais etapas necessárias para a aplicação da metodologia de mineração de dados em imagens de sensoriamento estão:

a. Pré-processamento;
b. Correção atmosférica;
c. Extração de dados;
d. Classificação;
e. Reconhecimento de padrões.

a) Pré-processamento de imagens de sensoriamento remoto

De acordo com Florenzano (2002), antes de gerar produtos como mapas e cartas-imagem, a partir de imagens adquiridas por satélites de observação da terra, é importante realizar um tratamento preliminar nos dados, com a finalidade de reduzir distorções gemétricas e radiométricas existentes na imagem.

Segundo Crósta (2002) quando o objetivo da análise das imagens é a produção de mapas, utiliza-se a correção geométrica para reduzir erros sistemáticos gerados no momento da aquisição das imagens pelo satélite. Além disso essa correção permite uma melhor interpretação da imagem enfatizando caracaterísticas de interesse para uma aplicação específica.

A correção geométrica se faz usualmente através da aplicação de um polinômio, cujos coeficientes são determinados a partir da relação entre coordenadas de imagem e coordenadas geográficas por meio de pontos de controle. Os pontos de controle são feições de fácil identificação (tal como confluência de rios e cruzamentos de estradas) tanto na imagem quanto num mapa.

Outro tipo de correção que também faz parte da fase de pré-processamento é a correção radiométrica. Trata-se de uma pratica pouco corrente quando o objetivo é a produção de mapas, que

Digitized by Google

envolve complexos algoritmos de processamento com entrada de dados de parâmetros específicos das condições atmosféricas e do sensor.

b) Correção atmosférica de dados de sensoriamento remoto

A atividade de sensoriamento remoto fundamenta-se na identificação e classificação de alvos, onde a radiação eletromagnética REM, percebida pelo sensor é utilizada para inferir as características dos mesmos. Todavia a atmosfera interage com a REM causando significativas alterações no fluxo de transporte da radiação solar por meio dos fenômenos de espalhamento, absorção e refração da energia eletromagnética.

De acordo com Liou (1980), isto ocorre quando a partícula atmosférica recebe a energia de uma onda eletromagnética incidente e reirradia a mesma em direções diferentes da original. É provocado pelos aerossóis, que são pequenas partículas finas sólidas ou líquidas em suspensão em um meio gasoso. Já a absorção consiste na transformação da energia eletromagnética em outra forma de energia que por sua vez é ocasionada, em geral, pelo vapor d'água, ozônio, oxigênio e dióxido de carbono.

Juntos, estes efeitos promovem a diminuição dos valores possíveis que seriam registrados pelo sensor, diminuição do contraste entre superfícies adjacentes e alteração do brilho de cada ponto da imagem (ZULLO Jr., 1994 *apud* NASCIMENTO, 2006). Assim a correção atmosférica torna-se um processo imprescindível, de modo que permite a remoção ou diminuição das influencias atmosféricas representando uma das fases mais importantes dentro do processamento digital de imagens de satélite (ROSA, 2001 *apud* NASCIMENTO, 2006).

Nascimento, 2006 destaca também que para a realização das correções é necessário o conhecimento das condições da atmosfera durante a aquisição de cada elemento da imagem (pixel). Somente assim é possível recuperar, com modelos matemáticos, o percurso da radiação e reconstruir as características espectrais da

Digitized by Google

superfície terrestre, sem falhas, considerando-se a atmosfera como um sistema altamente dinâmico e complexo que varia em escalas mínimas, visto que os constituintes atmosféricos apresentam elevadas variações.

c) Extrair informações de imagens de satélite

Segundo Meneses e Almeida (2012), para extrair informações das imagens é necessário estabelecer métodos de análise com regras claras e lógicas, para que possam ser replicadas por qualquer interessado na interpretação do significado do dado. Essas regras estabelecem critérios para a análise dos elementos que compõem a paisagem, utilizando-se das propriedades da cor, tonalidade, textura, estrutura, sombras e homologia que são específicas de cada classe de alvo.

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, por exemplo, contêm informações importantes sobre o arranjo estrutural das superfícies e suas relações com o entorno (HARALICK *et al.*, 1973).

Segundo Batista (2002), a textura é definida como uma combinação entre a magnitude e a frequência da variação tonal numa imagem e é visualmente interpretada por uma repetição tonal de pixels em uma determinada região desta imagem. Diferentemente dos demais atributos relativos às propriedades espectrais dos objetos, a textura contém informações sobre a distribuição espacial dos objetos.

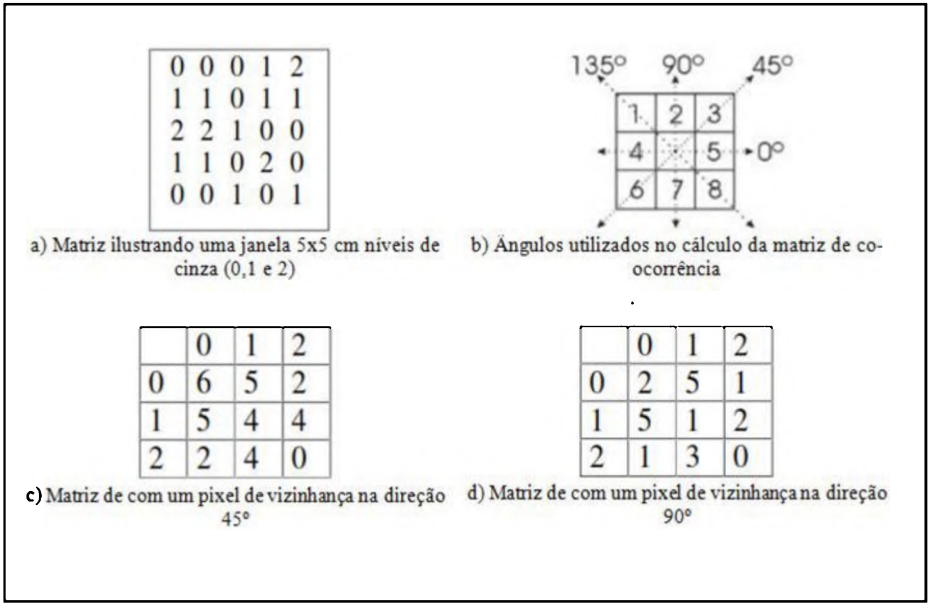
Os atributos de texturas são geralmente divididos em três grupos: os estruturais e os estatísticos (HE *et al.*, 1987 *apud* BATISTA, 2002).

Os atributos estruturais consideram a textura como uma repetição de padrões primitivos básicos, não considerando a distribuição espacial dos níveis de cinza. Os métodos estatísticos são propostas mais usuais para a extração de atributos de textura e fo-

Digitized by Google

ram introduzidos por Haralick *et al.*, (1973), com o objetivo de explorar a dependência espacial dos níveis de cinza da imagem para montar a matriz que melhor representa a variação tonal. A matriz de co-ocorrência ilustrada na Figura 3a, é definida para uma orientação e uma distância, conforme a Figura 3b, entre os pixels da imagem. Nas Figuras 3c e Figura 3d são ilustradas as matrizes de co-ocorrência com 1 pixels na vizinhança nas direções 45º e 90º, respectivamente.

Figura 3 – Ilustração da construção da matriz de co-ocorrência



Fonte: VASCONCELOS (2010)

Haralick (1973) *apud* Batista (2006) sugere que os principais atributo 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. Na Tabela 4 são descritas as definições de alguns dos atributos de textura utilizados neste trabalho. O conjunto completo de atributos e as respectivas definições podem ser encontradas em Batista (2002).

Digitized by Google

Tabela 4: Definição de alguns atributos de textura

Contraste: O atributo contraste diz respeito à diferença entre os mais altos e os mais baixos valores de um conjunto de pixels adjacentes.	$$c = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$$ Onde C representa o valor do contraste, i a identificação da linha, j é a identificação da coluna, $P_{i,j}$ é o valor normalizado para os pixels i e j e N é o número de linhas ou de colunas.
Entropia: O atributo entropia estima a aleatoriedade dos dados, assumindo valores baixos para imagens suaves e homogêneas.	$$e = \sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j})$$ Onde e representa o valor da entropia, i a identificação da linha, $P_{i,j}$ é o valor normalizado para os pixels i e j e N é o número de linhas ou de colunas.
Segundo Momento Angular: Este atributo avalia a uniformidade da textura e a suavidade local.	$$asm = \sum_{i,j=0}^{N-1} P_{i,j}^{\ 2}$$ Onde Onde asm representa o momento angular de segunda ordem, $P_{i,j}$ é o valor normalizado para os pixels i e j e N é o número de linhas ou de colunas.

Fonte: BATISTA (2002)

Além da experiência do interprete da imagem, fatores como o número de bandas que deve ser analisado, a extensão da área de estudo ou a análise conjunta de várias imagens multitemporais tornam bastante demorado o trabalho de interpretação. Como reposta a esses problemas surgiram as técnicas de classificação digital de imagens, que automatizam o processo de extração das informações das imagens eliminando a subjetividade da interpretação humana e reduzindo a dificuldade empregado pelo interprete.

d) Classificação de imagens de sensoriamento remoto

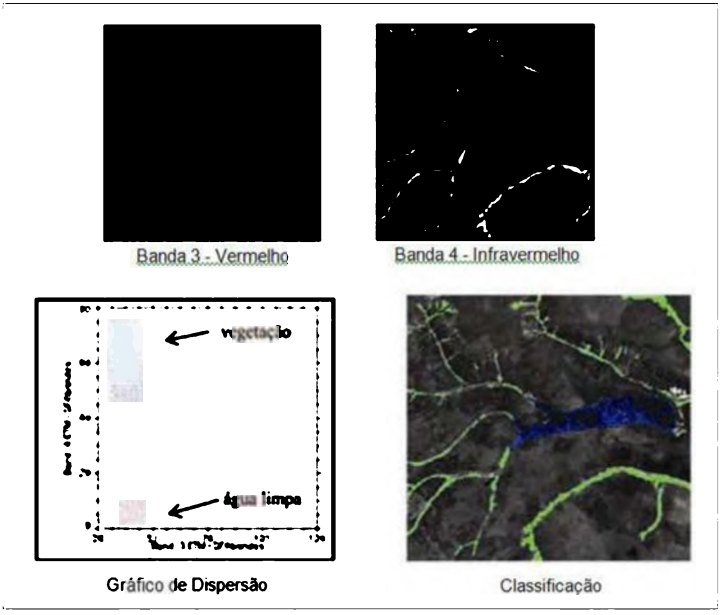
Classificação de imagens multiespectrais é o processo de associar os pixels da imagem a um número finito de classes individuais que representam os objetivos do mundo real, com base nos seus

Digitized by Google

valores digitais. Se um pixel satisfaz um determinado conjunto de critérios, normalmente avaliados por medidas estatísticas, o pixel é atribuído á classe que melhor satisfaz os critérios estabelecidos (MENESES e ALMEIDA, 2012).

Desse modo o resultado final de um processo de classificação é uma imagem digital que constitui um mapa de "pixels" classificados, representados por símbolos, gráficos ou cores, conforme representado na Figura 4, onde cada cor ou símbolo pode ser associado a uma classe (tipo de solo ou vegetação) definida pelo usuário.

Figura 4 - Classificação de imagens



Fonte: Meneses e Almeida (2012)

e) Reconhecimento de padrões

Segundo Menezes e Almeida (2012), reconhecimento de padrões é a ciência de encontrar padrões significativos nos dados da imagem que possam ser extraídos por operadores automatizados. O reconhecimento de padrões é realizado pelo cérebro humano que rotula certas texturas e cores em classes de alvos ou objetos.

Digitized by Google

Esse é um processo realizado por treinamento neurocerebral desenvolvido ao longo da experiência de vida. Em um computador, o reconhecimento de padrões em imagens pode ser feito de modo mais científico. Medições estatísticas são derivadas a partir das estatísticas espectrais dos pixels em uma imagem, a fim de que todos os pixels sejam destinados às suas mais prováveis classes, usando uma régua de decisão (MENESES e ALMEIDA, 2012).

Meneses e Almeida (2012) ressaltam que quem faz a delimitação dos diferentes polígonos de cada classe na imagem é o algoritmo de classificação digital, mas quem constrói a legenda do mapa temático continua sendo o analista. Aos valores de pixels que identificam uma classe específica de alvo ou objeto real da natureza, denomina-se padrões. Portanto a classificação automática de imagens é nada mais do que um processo de reconhecimento de padrões.

2.2 Descoberta de conhecimento em bancos de dados

Nas últimas décadas, em consequência dos avanços tecnológicos ocorridos em diversas áreas da ciência, pesquisadores e analistas experimentam um grande aumento no volume e variedade dos dados. O sensoriamento remoto orbital é um dos exemplos de tecnologia onde tem ocorrido um aumento exponencial da quantidade de dados sobre as características do planeta Terra, que são coletados pelos sensores. Os custos de armazenamentos decrescentes viabilizam ainda mais esse aumento na produção de dados. Nesse contexto explicita-se a necessidade de se desenvolver novas técnicas e ferramentas automatizadas que permitam determinar a relevância em meio aos grandes volumes de dados e criar valor a partir dessa relevância, isto é, transformar dados em informação útil e conhecimento (HAN e KANBER, 2006).

De acordo com FAYYAD *et al.*, (1996a), a descoberta de conhecimento em bancos de dados ou KDD (*Knowlegde Discovery in*

Digitized by Google

Databases), termo inglês mais utilizado na literatura, é o processo de identificar em dados, padrões que sejam válidos, novos (antes desconhecidos), potencialmente úteis e interpretáveis, visando melhorar o entendimento de um problema ou um procedimento de tomada de decisão. Estudando os conceitos acima individualmente, obtém-se:

- **Dados:** um conjunto ou subconjunto de dados F, como campos de um banco de dados, como por exemplo, uma coleção de n cadastros de clientes, contendo idade, profissão, renda ou, no contexto deste trabalho uma coleção de pixels referentes a áreas com florestas plantadas em uma imagem de sensoriamento remoto e seus respectivos valores de reflectância, segundo a banda do espectro.
- **Padrão:** expressão E em uma linguagem L descrevendo fatos em um subconjunto F_E de F. É dito um padrão se é mais simples do que a enumeração de todos os fatos com $F_{E.}$ Por exemplo, o padrão "área com declividade < r% então área considerada tem vocação florestal" seria aplicado para uma escolha apropriada de r. Fazendo um paralelo a este trabalho, "se a energia refletida do pixel de uma área com reflorestamento > w, então o alvo não representa o alvo de interesse" poderia ser aplicado para uma escolha conveniente de w.
- **Processo**: geralmente em KDD, processo geral de descoberta do conhecimento composto por várias etapas que envolvem preparação dos dados, avaliação de conhecimento, pesquisa de padrões e refinamento envolvendo iteração e modificação.
- **Validade**: Os padrões descobertos devem ser validos em novos dados com algum grau de certeza. Uma medida de interesse é uma função C mapeando expressões em L para um espaço de medidas Mc Por exemplo, se uma regra de classificação restringe-se a apenas uma classe de solo, então a medida de certeza diminuiria conforme o

Digitized by Google

modelo de classificação fosse aplicado a novas imagens contendo classes de solo diferentes daquelas utilizadas inicialmente.

- **Novo**: em geral, assume-se que "novidade" pode ser medida por uma função *N*(A, B), que pode ser uma função booleana ou uma medida que expresse grau de "novidade" ou "surpresa". Como exemplo de um caso que não é novo: sejam *A*= "vegetação" e *B*= "alta reflectância na faixa do infravermelho" então *N*(A, B) = 0 ou *N*(A, B) = false. No entanto, sejam *A* = "reflectância do pixel de área cultivada com reflorestamentos" e *B* = " índice de área foliar ", então *N*(E, F) = 0,85 ou *N*(E, F) = true.
- **Potencialmente útil:** padrões devem, essencialmente, levar a alguma atitude prática, conforme medida gerada por alguma função de utilidade. Por exemplo regras obtidas no processo de classificação podem ser aplicada na identificação de novas áreas de reflorestamento e, assim, orientar o trabalho de instituições interessadas no inventario de áreas com reflorestamentos ou diagnosticar o setor.
- **Compreensível**: Um dos objetivos de KDD é tornar padrões compreensíveis para humanos, com a finalidade de promover uma melhor compreensão dos próprios dados. Embora a tarefa de medir compreensibilidade seja complexa, uma das formas frequentemente utilizadas é a medida de intuitividade. Entende-se por compreensão dos dados o grau de simplicidade da representação destes, bem como do nível de detalhe alto o suficiente para que estes sejam compreendidos. Por exemplo: o log de um servidor Web não é uma representação compreensível; estatísticas extraídas deste log, tais como o número total de acessos ou a qualificação dos tipos de usuários, fornecem informação num formato mais intuitivo e humanamente compreensível.

Digitized by Google

O KDD também é entendido como uma intersecção de áreas de pesquisa afins como aprendizado de máquina, reconhecimento de padrões, estatística, visualização de dados e computação. Dessa maneira os processos e estudos nesta área continuam em constante evolução segundo Fayad *et al.*, (1996b).

Segundo Rezende *et al.*, (2002) o KDD se fundamenta em técnicas conhecidas de aprendizado de máquina, de reconhecimento de padrões e de estatística para encontrar padrões nos dados. Este processo exige a concentração de três atores de habilidades distintas, que trabalham em cooperação, e o seu sucesso depende, em parte, dessa cooperação. Os atores do processo podem ser divididos em três classes: o Usuário, o Especialista do domínio e o Analista de Dados, sendo que:

- **Especialista do domínio:** é quem conhece o tema que será estudado e possui o domínio de aplicação (normalmente um pesquisador ou profissional com larga experiência) e deve fornecer apoio para a execução do processo.
- **Analista de dados:** pessoa que executa o processo de verificação de hipóteses levantadas pelo Especialista de Domínio (normalmente em conjunto com o Usuário). Este usuário deve conhecer a fundo as etapas que compõem o processo.
- **Usuário final:** é quem vai utilizar o conhecimento obtido como auxílio em um processo de tomada de decisão e caso necessário reformular hipóteses que serão novamente testadas até que o problema levantado seja solucionado.

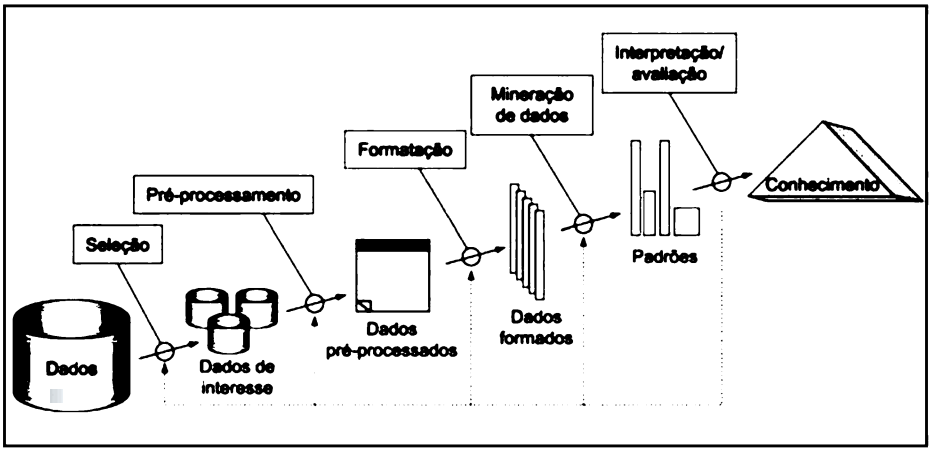
2.3 Fases da descoberta de conhecimento em bancos de dados (KDD)

Fayyad *et al.*, (1996) destaca que o processo de KDD é interativo, iterativo, cognitivo e exploratório, composto por várias etapas. A Figura 5 apresenta esse processo, onde a participação

Digitized by Google

do analista (profissional especializado em análise de dados) ocorre diversas vezes.

Figura 5- Fases e processos em KDD



Fonte: Fayyad *et al.*, (1996a)

As etapas do processo representado na Figura 5 não precisam ser seguidas obrigatoriamente na ordem descrita. Considerando que o processo de descoberta de conhecimento é iterativo e exploratório como visto anteriormente, algumas de suas etapas podem ser executadas novamente dependendo do resultado de etapas anteriores.

Após a determinação do conhecimento relevante que se pretende descobrir, o que sugere uma compreensão do domínio da aplicação, bem como dos objetivos do usuário final do processo, iniciam-se as fases do processo de KDD, conforme os itens a seguir:

a) Seleção de dados

A realização dessa etapa de maneira apropriada pode viabilizar a extração de conhecimento útil dos dados. Por outro lado, a não realização das etapas constituintes do processo de descoberta do conhecimento em bancos de dados tende à descoberta de padrões não reais ou sem sentido prático (AGRAWAL *et al.*, 1996 *apud* NONATO, 2013).

Digitized by Google

Como ilustração da etapa de preparação de dados aplicada ao sensoriamento remoto pode-se citar a produção de composições coloridas utilizadas na classificação de dados, pois são formadas por varias imagens de diferentes bandas do espectro eletromagnético ou, de imagens índice de vegetação. Para facilitar a compreensão, para uma dada imagem em uma determinada banda, podem ser extraídos três campos contendo: latitude, longitude e número digital (DN), onde cada registro corresponde ao vetor de atributos de um dado pixel. Como suposição, uma composição resultante da sobreposição de imagens em três bandas distintas, teria um vetor de atributos de tamanho três, ou seja, um registro com três campos onde cada campo é interpretado como uma variável de estudo.

b) Pré-processamento dos dados e limpeza

Consiste na remoção de ruídos e desvios (se possível e apropriado), a coleta da informação necessária para modelar ou estimar ruído, a decisão de como proceder com campos de dados ou atributos faltantes, e a formatação de dados de forma a adequá-los às técnicas de mineração.

Segundo MANNILA (1996), a etapa de pré-processamento é considerada de grande complexidade e pode ocupar 80 % de todo o tempo de processo e precisa ser realizada por especialista com domínio do problema em questão.

A limpeza dos dados envolve a verificação da consistência nos dados: como, por exemplo, a existência de registros incompletos ou de valores errados. Essa verificação exige a tomada de decisão sobre como proceder em casos onde existem atributos incompletos.

c) Transformação dos dados

Também conhecida como fase Redução ou Formatação dos dados, a Transformação consiste na seleção de características úteis ou relevantes para representar os dados (escala, orientação), sempre dependendo do objetivo a ser alcançado, visando à redu-

Digitized by Google

ção do número de variáveis e/ou instâncias a serem consideradas para o conjunto de dados Esta fase deve levar em consideração peculiaridades do algoritmo que será utilizado na mineração de dados. Entre as transformações utilizadas estão: normalização de variáveis quantitativas e transformação de variáveis qualitativas em quantitativas (Fayyad *et al.*, 1996).

d) Mineração de dados

Definição das tarefas de mineração considerando os objetivos/metas do processo (classificação, regressão, agrupamento); 2) Definição dos algoritmos, o que envolve a escolha do modelo, parâmetros e formas de execução; 3) busca de padrões de interesse utilizando os algoritmos e parâmetros selecionados.

e) Interpretação dos padrões e consolidação do conhecimento descoberto

Envolve a avaliação dos padrões e regras encontrados para confirmar se estes representam conhecimento novo em situações práticas. Posteriormente, pode-se realizar a documentação e compartilhamento com partes interessadas neste conhecimento, considerando possíveis conflitos (Fayyad *et al.*, 1996).

Para exemplificar possíveis aplicações de KDD na área de sensoriamento remoto, são apresentados a seguir casos onde a descoberta de conhecimento em bancos de dados foi utilizada.

Silva (2006) apresenta uma proposta para extrair padrões de mudança a partir de imagens de sensoriamento remoto utilizando conceitos de processamento digital de imagens, mineração de dados e ecologia da paisagem. O autor destaca que diante da problemática sócio-ambiental advinda do rápido desflorestamento da Amazônia, o uso da estratégia de Descoberta de Conhecimento em Bancos de Dados (KDD), auxiliou na compreensão de processos de mudança de uso do solo, bem como na identificação e análise de processos de desflorestamento em áreas amazônicas.

Celinski (2008) trabalhou a discriminação de classes de

Digitized by Google

cobertura do solo em imagens de sensoriamento remoto do satélite CBERS-2 por meio do classificador Árvore de Decisão. O estudo incluiu a avaliação e a verificação da acurácia da metodologia proposta comparativamente ao Classificador de Máxima Verossimilhança (MAXVER). Foi utilizado um conjunto de (16) atributos a partir das imagens, composto por: bandas do sensor CCD (1, 2, 3, 4), índices de vegetação (CTVI, DVI, GEMI, NDVI, SR, SAVI, TVI), componentes de mistura (solo, sombra, vegetação) e os dois primeiros componentes principais. A acurácia da classificação foi avaliada por meio da matriz de erros de classificação e do coeficiente kappa.

2.4 Mineração de dados

Como visto anteriormente, o KDD é composto por várias fases, e destaca-se que todas as fases são importantes. Usualmente a fase de análise é a mais complexa, e compreende, entre outras, a etapa de Mineração de dados frequentemente confundida na literatura como o próprio processo de Descoberta de Conhecimento em Bancos de Dados que é um processo mais abrangente. Esse possível equívoco ocorre porque a etapa de mineração de dados é uma das etapas essenciais do processo. Dessa maneira, a mineração de dados é uma etapa dentro de um processo mais geral (HAN e KAMBER, 2006).

Segundo Fayyad, (1996a), KDD refere-se ao processo global de descoberta do conhecimento a partir de dados, e a atividade de mineração é um passo particular deste processo. Dentro deste contexto a mineração de dados deve ser entendida como a aplicação de algoritmos específicos para extrair padrões dos dados. Os passos adicionais no processo, como preparação, seleção, limpeza e consolidação também são importantes, pois garantem a qualidade e a utilidade do conhecimento adquirido através dos dados.

Digitized by Google

Durante o processo de Descoberta de Conhecimento, é importante distinguir o que é uma tarefa e o que é uma técnica de mineração de dados. Uma tarefa de mineração de dados consiste na especificação de que tipo de regularidades ou categorias de padrões estamos interessados em encontrar nos dados, como por exemplo, um padrão da frequência de chuva em determinada região, a classificação de uma área com vocação florestal a partir das condições edafoclimáticas. Já a técnica de mineração de dados consiste na especificação de metodologias que permitam encontrar os padrões de interesse, isto é, a realização de uma tarefa especificada anteriormente (AMO, 2004 *apud* NONATO, 2013).

De acordo com MEIRA (2008) não existe a melhor técnica de mineração de dados já que cada uma possui pontos positivos ou negativos. A escolha de uma determinada técnica precisa de um estudo detalhado do problema em questão e a definição de qual estratégia de descoberta melhor se adéqua. No caso particular deste trabalho, a principal tarefa é a classificação com ênfase em árvores de decisão.

2.4.1 Principais tarefas de mineração de dados

As tarefas de mineração de dados mais comuns são apresentadas na Figura 6. A tarefa de predição é similar às tarefas de classificação e regressão e envolve o uso de variáveis com valores conhecidos para predizer um valor desconhecido ou futuro de outra variável. Já a tarefa de descrição é muito utilizada durante a análise exploratória de dados, para comprovar a influência de certas variáveis e oferecer uma possível interpretação para os resultados obtidos. Esses objetivos podem ser alcançados por meio de vários tipos de tarefas. A escolha de uma ou mais tarefas, por sua vez, depende do problema em questão (HAN e KANBER, 2006; FAYYAD *et al.*, 1996b).

Digitized by Google

Figura 6 - Tarefas de mineração de dados

Fonte: REZENDE et al. (2002)

Dentre as principais técnicas utilizadas em mineração de dados, têm-se técnicas estatísticas, técnicas de aprendizado de máquina e técnicas baseadas em crescimento poda e validação (HAN e KANMER, 2006; AMO, 2004 *apud* NONATO, 2013). A seguir, são brevemente descritas as principais tarefas de mineração de dados:

a) Classificação

É o processo de encontrar um ou mais modelos (funções) que classifiquem um registro como pertencente a uma classe dentre algumas pré-definidas. Esse modelo têm o propósito de utilizá-los para predizer classes de objetos que ainda não foram classificados. Como exemplo, no âmbito deste trabalho, pode-se ser mencionada a classificação de uma coleção de pixels de uma imagem como uma cultura agrícola, solo exposto, água ou vegetação nativa.

b) Regressão

Aplicada em situações onde o interesse do usuário é estimar valores desconhecidos em uma variável contínua. Por exemplo: densidade do biodiesel de diferentes matérias primas. Neste caso, a tarefa de mineração é denominada Regressão e uma das técnicas que podem ser empregadas para atingir este fim é a análise de regressão.

Digitized by Google

A análise de regressão consiste em descobrir uma função que mapeie um item de dados para uma variável de predição real estimada. Por exemplo, no âmbito deste trabalho, poderia ser de interesse modelar o desenvolvimento de reflorestamentos a partir da intensidade de reflectância registrada em um dado pixel da imagem.

Embora o método de regressão seja bastante difundido em estudos de estatística, é pouco utilizado em problemas de mineração de dados, pois na vida real problemas envolvendo classificação são mais comuns.

c) Regras de associação

Associação: uma regra de associação é um padrão da forma $X \rightarrow Y$, onde X e Y são conjuntos de valores, isto é, o quanto a presença de um conjunto de itens nos registros de itens em um banco de dados implica na presença de algum outro conjunto de itens nos mesmos registros. Por exemplo, consideremos o caso de escoamento de produção. O padrão "Produção próxima do terminal de transporte então menor o gasto com frete " representa uma regra de associação que reflete um padrão de custos de transporte. Nesse sentido, descobrir regras de associação entre a localização das áreas de cultivo e a proximidade das linhas de transporte pode ser útil para melhorar a organização da produção e reduzir os custos de transporte.

d) Agrupamento

Está tarefa difere da classificação, pois a análise de cluster (agrupamentos) trabalha sobre dados onde as denominações das classes não estão definidas, isto é, não necessita que os registros sejam previamente categorizados (aprendizado não-supervisionado). A tarefa consiste em particionar uma população heterogenia em vários subgrupos mais homogêneos. Os dados são agrupados de acordo com as similaridades, o que a diferencia da tarefa de Classificação.

Digitized by Google

Por exemplo, a análise de clusters poderia ser aplicada sobre uma base de dados de imóveis rurais a fim de identificar grupos homogêneos de cultivos, tais como: imóveis localizados na região Sul cultivam a cultura A, imóveis localizados na região Norte cultivam a cultura B, etc. No contexto deste trabalho recomenda-se a execução de procedimento de classificação não supervisionada, como a análise de cluster na etapa inicial do processo de construção de mapas temáticos.

e) Análise de desvios (*Outliers*)

A tarefa de análise de desvios tem por objetivo descobrir em um banco de dados valores que apresentem o comportamento adverso. O termo mais utilizado para denominar estes registros incomums é "*outliers*". Muitos métodos de mineração desprezam estas exceções (*outliers*). Entretanto em algumas aplicações, como as que estudam risco ou modelagem de eventos raros, estes *outliers* podem ser mais interessantes do que eventos que ocorrem cotidianamente. Por exemplo, pode-se detectar variações das condições climáticas durante o ciclo de cultivo fora de seu padrão habitual e melhorar a gestão de riscos associados a cultivos agroenergéticos.

Dentre as diversas metodologias de classificação de dados conhecidas na literatura científica e também utilizadas na prática, algumas são utilizadas com maior frequência pela simplicidade, viabilidade e clareza dos resultados. Pode-se citar entre os algoritmos classificadores tradicionais: os métodos de máxima-verossimilhança, as técnicas de regressão, o discriminante linear de Fisher, os classificadores bayesianos, as metodologias baseadas em distância como a euclidiana, as redes neurais, a lógica fuzzy, entre outras.

Uma maneira de avaliar o que torna alguns classificadores mais vantajosos do que outros no processo de mineração de dados, reside no fato de que em geral, no processo de KDD e consequentemente de mineração de dados, evita-se o emprego de suposições preliminares sobre os dados. A título de ilustração, enquanto no modelo baseado em máxima-verossimilhança há a exigência de

Digitized by Google

uma distribuição estatística conhecida e resultados do classificador baseados na sua acurácia, no modelo de árvore de decisão, não há essa exigência e no contexto de sensoriamento remoto pode realizar classificação de imagens de forma relativamente rápida, além de trazer como resultado um conjunto de regras de classificação interpretáveis pelo usuário (Lillesand e Kiefer 1994).

Fazendo um comparativo entre os classificadores não-paramétricos, as redes neurais artificiais, por exemplo, embora possam ter precisão elevada, são relativamente difíceis de compreender quando comparadas com árvores de decisão. O analista parte da suposição que o relacionamento entre as variáveis explanatórias e o atributo alvo são definidos por uma função definida a priori. Feita esta suposição, o objetivo do analista passa a ser encontrar os parâmetros que reproduzam o melhor ajuste daquela função aos dados em estudo. Geralmente os modelos baseados em redes neurais, são chamados de modelos de "caixa preta" devido à dificuldade encontrada na extração de regras justificáveis da decisão tomada pela rede. (FAYAAD *et al.*, 1996a).

2.5 Árvores de decisão

De acordo com Monard e Baranauskas (2002) *apud* Meira (2008), as regras de classificação extraídas após a modelagem por árvore de decisão "*são consideradas técnicas de aprendizado orientadas a conhecimento, em que o interesse principal consiste em obter descrições simbólicas que sejam de compreensão fácil e utilização por meio de modelos mentais*". (p.10). Dentre as muitas metodologias de classificação de dados existentes, as árvores de decisão são modelos não-paramétricos adequados ao processo de KDD, pois dão ênfase especial no descobrimento de padrões compreensíveis que possam ser interpretados como conhecimento útil ou interessante.

Com características peculiares, uma árvore de decisão tem a função de particionar recursivamente um conjunto de treinamen-

Digitized by Google

to até que cada subconjunto obtido contenha casos de uma única classe. Para atingir esta meta, o algoritmo escolhido para a árvore de decisão examina e compara a distribuição de classes durante a construção da classe.

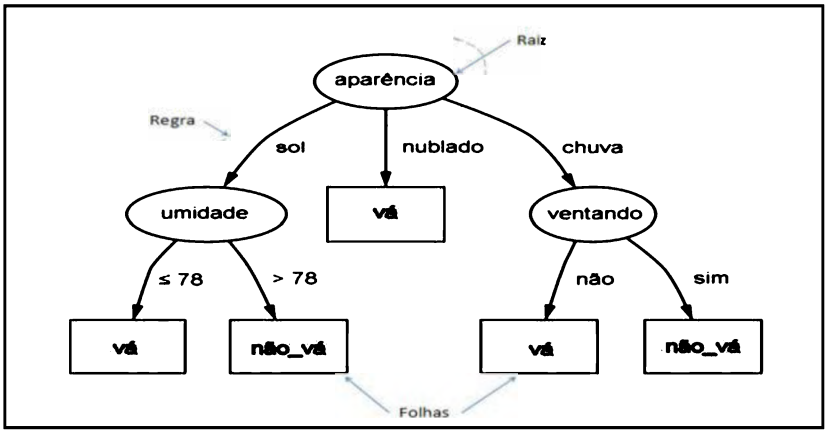
As árvores de decisão são formadas de nodos, que representam os atributos, de arcos provenientes destes nodos, que recebem os valores possíveis para esses atributos, e de nodos folha, que representam as classes distintas de um conjunto de treinamento. Desse modo, a classificação por árvores de decisão pode ser utilizada para classificar todos os objetos do conjunto de dados da entrada. Os resultados obtidos após a construção de uma árvore de decisão, são dados organizados de maneira compacta, com a árvore podendo ser utilizada para classificar novos casos (BRAZDIL, 2005; INGARGIOLA, 2010 *apud* MEIRA, 2008).

Segundo Han e Kamber (2006), a construção de uma árvore de decisão é baseada num modelo *Top-Dowm* representado graficamente por nós e ramos, isto é, do nó raiz em direção as folhas. O nó raiz é o primeiro nó da árvore, no topo da estrutura. Os nós internos, incluído o nó raiz, são nós de decisão. Cada nó contém um teste relacionado a um ou mais atributos e os resultados deste teste formam os ramos da árvore. Geralmente, o teste em um nó compara o valor de um atributo com um valor constante. No entanto, algumas árvores podem comparar dois atributos entre si, ou utilizar uma função envolvendo um ou mais atributos (WITTEN e FRANK, 2005).

De modo geral os classificadores baseados em árvore de decisão, assim como outros, fundamentam-se na filosofia da divisão do problema em vários subproblemas de menores dimensões, até que uma solução para cada um dos problemas mais simples seja encontrada. Como exemplo, a Figura 7, apresenta uma aplicação de um modelo de árvore de decisão hipotética com o objetivo de recomendar, ou não, uma viagem, sabidas as condições do tempo.

Digitized by Google

Figura 7 - Exemplo de uma árvore de decisão



Fonte: Adaptado de MONARD e BARANAUSKAS (2002) apud MEIRA (2008)

O atributo alvo de interesse possui duas classes: **vá viajar** ou **não vá viajar**.

- **Nó:** São todos os itens que não aparecem na árvore;
- **Folhas:** São nós que não tem filhos, os últimos itens da árvore;
- **Filhos:** São os itens logo abaixo da raiz;
- **Raiz:** é o item topo da árvore.

Conforme ilustrado na Figura 7 um caminho é traçado iniciando do nó raiz, percorrendo os ramos, de acordo com os resultados do teste, até chegar a uma só folha, que representa a classe de predição do exemplo (HAN e KAMBER, 2006). Assim em uma árvore de decisão tem-se que cada nó interno é 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 ate a folha.

2.5.1 Regras de classificação

Analisando a estrutura da árvore de decisão da Figura 7 é

Digitized by Google

possível perceber que o conhecimento representado em árvores de decisão pode ser derivado em regras de classificação da forma SE-ENTÃO. Uma regra é criada para cada caminho entre a raiz e um nó folha. As regras são escritas considerando o trajeto do nó raiz até uma folha da árvore. Os antecedentes da regra são formados pelos atributos preditivos que aparecem ao longo do caminho percorrido, testando os valores que os definem e os consequentes são formados pelo atributo classe. Essas regras podem ser mais difíceis de compreender, principalmente se a árvore de decisão for muito grande, sendo que nestes casos é mais fácil trabalhar com as regras (HAN e KAMBER, 2006). A seguir estão as regras de classificação da árvore de decisão ilustrada na Figura 6.

- SE aparência = sol E umidade <= 78 ENTÃO classe = vá
- SE aparência = sol E umidade > 78 ENTÃO classe = não_vá
- SE aparência = nublado ENTÃO classe = vá
- SE aparência = chuva E ventando ENTÃO classe = vá
- SE aparência = chuva E ventando ENTÃO classe = não_vá

Uma peculiaridade das regras extraídas de uma árvore de decisão é a não ambiguidade e a disjunção das mesmas, deste modo, a ordem com que são executadas não importa. Quando as regras são muito difíceis de entender, é possível remover testes redundantes (MONARD e BARANAUSKAS, 2002 apud MEIRA, 2008; WITTEN e FRANK, 2005).

2.5.2 Avaliação da Classificação

As metodologias de avaliação de classificadores têm como princípio fundamental a mensuração da acurácia dos resultados obtidos após a aplicação destes a novos conjuntos de dados. Espera-se que o classificador alcance uma taxa de 100% de acerto nos novos dados e uma taxa de 0% de erro. Como isso dificilmente acontece, o analista deve buscar configurações capazes de alcanças as maiores taxas de acertos e menores taxas de erro possível.

Outro ponto importante, principalmente no contexto desse

Digitized by Google

trabalho, é que um teste completo de acurácia de um mapa de classificação exigiria a verificação da classe de cada pixel no mapa, o que raramente é possível. Dessa forma, áreas de teste representativas devem ser utilizadas para se estimar a acurácia do mapa, que é quase sempre estimada pela divisão do número de pixels de teste classificados corretamente na classe pelo número total de pixels na classe (SCHOWENGERT, 1983 *apud* MENESES e ALMEIDA, 2012).

Em geral, as taxas de erros e acertos devem ser obtidas após a aplicação do modelo em um conjunto de dados diferente daquele utilizado na fase de treinamento. Isto é importante para evitar um ajuste específico (*overffiting*) do modelo aos dados de treinamento. A principal consequência do *overffiting* é uma alta taxa de erro na aplicação do modelo em novos dados.

Segundo Han e Kamber (2006), as abordagens mais utilizadas para solucionar esse problema são *split-sample*, método tradicional que 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.

Outro método bastante utilizado é o *cross-validation* - validação cruzada, utilizado principalmente quando a quantidade de dados para dividir entre treinamento e teste é limitada (WITTEN e FRANK, 2005). Na validação cruzada, os dados são aleatoriamente divididos em k partições mutuamente exclusivas (*folds*) de tamanho aproximadamente igual.

O modelo é então treinado *k* vezes, onde cada vez uma partição é utilizada para teste, o as demais juntas são utilizadas para treinamento. A taxa de erro, ao final, é calculada como a média das taxas de erro obtida em cada uma das partições de teste. A vantagem da validação cruzada é utilizar cada um dos exemplos para treinamento e para 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 chama-

Digitized by Google

da matriz de confusão representada na Tabela 4 e amplamente utilizada em análises estatísticas de concordância.

Considerando ainda a matriz de confusão da Tabela 5, que ilustra o caso relativo a um problema com duas classes, denominadas C_+ (classe positiva) e C_- (classe negativa), as quatro possibilidades de acertos e de erros do classificador podem ser identificadas como:

- Verdadeiros positivos (VP), casos onde os objetos pertencem C_- e foram preditos como pertencentes a essa mesma classe;
- Falsos negativos (FN) casos onde os objetos pertencem à classe C_+ e foram preditos como pertencente à classe C_-;
- Verdadeiros negativos (VN), casos onde os objetos pertencem à classe C_- e foram preditos como pertencentes a essa mesma classe;
- Falsos preditos (FP), casos onde os obtetos pertencem à classe C_- e foram preditos como pertencente à classe C_+.

Tabela 5 – Matriz de confusão para a classificação com duas classes

		Predita	
		C_+	C_-
Verdadeira	C_+	VP	FN
	C_-	FP	VN

Fonte: MEIRA (2008)

Além da própria acurácia e da taxa de erro, outras medidas podem ser derivadas da matriz de confusão, (Tabela 5), tais como (MONARD e BARANAUSKAS, 2002 *apud* MEIRA, 2008):

Acurácia (taxa de acerto) mede a proporção do número total de predições que foram classificadas corretamente. Nessa definição, n é o número total de objetos classificados.

Digitized by Google

$$acc = \frac{VP + VN}{n} \quad 3.1$$

A **sensibilidade** é definida como a proporção de casos positivos que foram corretamente classificados.

$$sens = \frac{VP}{VP + VN} \quad 3.2$$

Especificidade é definida como a proporção de casos negativos que foram classificados corretamente.

$$spec = \frac{VN}{VN + FP} \quad 3.3$$

Confiabilidade positiva ou precisão é definida como a proporção de casos positivos que de fato eram positivos.

$$prel = \frac{VP}{VP + FP} \quad 3.4$$

A **confiabilidade negativa** é definida como a proporção de casos negativos que foram classificados como positivos.

$$nrel = \frac{VN}{VN + FN} \quad 3.5$$

2.5.3 Estatística Kappa de Cohem e a Análise de Concordância

Além das medidas já descritas anteriormente, para saber se a classificação de um objeto é confiável, é necessário ter este objeto discriminado ou classificado várias vezes, por exemplo, por mais de um julgador ou classificador. Segundo Cohen (1960), 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

Digitized by Google

kappa é uma medida de concordância e indica como cada classificação difere de uma classificação aleatória dos tipos de cobertura, de acordo com Lillesand e Kiefer (1994). O valor máximo de desta medida de concordância é 1, onde este valor representa total concordância e os valores próximos a 0, representam nenhuma concordância, ou a concordância foi exatamente a esperada pelo acaso. A Tabela 6 exibe a interpretação:

Tabela 6 - Classificação qualitativa quanto ao valor da estatística *Kappa*

Estatística *Kappa*	Qualidade da classificação
K = 0,0	Péssima
0,0 < 0,2	Ruim
0,2 < 0,4	Razoável
0,4 < 0,6	Boa
0,6 < 0,8	Muito boa
0,8 < 1,0	Excelente

Fonte: CELINSK (2008)

Para avaliar se a concordância é razoável, pode ser necessário um teste estatístico de hipótese para avaliar a significância da estatística *kappa*. Neste caso a hipótese a ser testada seria a de que a estatística *kappa* é igual a 0, o que indicaria concordância nula, ou se ele é maior do que zero, indicando uma concordância maior do que o acaso. Um *kappa* com valor negativo, que não possui interpretação aceitável, pode resultar num paradoxal nível crítico (valor de *p*) maior do que 1, (LANDIS e KICH, 1977).

No caso de rejeição da hipótese nula (*Kappa* = 0) temos a indicação de que a medida de concordância é significativamente maior do que zero, o que indicaria que existe alguma concordância. Todavia isso não significa que a concordância seja alta, sendo, portanto necessária à avaliação por parte do pesquisador da adequação da medida obtida. Landis e Kick (1977) apresentam uma Tabela que permite avaliar o valor do coeficiente *kappa* obtido a partir de uma classificação.

Digitized by Google

CAPÍTULO 3: PANORAMA DAS FLORESTAS PLANTADAS NO MUNDO, NO BRASIL E NO TOCANTINS

3.1 Panorama do setor florestal no mundo.

A população mundial aumenta continuamente, fato que promove uma maior demanda por alimentos, fibra e energia. Em vários países em desenvolvimento como o Brasil, não apenas cresce o número de habitantes, como também o poder de compra das pessoas, que consomem cada vez mais produtos essenciais e supérfluos. Dessa maneira, é natural que a demanda por matéria-prima cresça aceleradamente. A madeira é uma entre as matérias primas, presente em todos os lares, ruas, comércios, na forma de móveis, papel e até mesmo aço (CIPRIANI, 2013).

Estima-se que em 2010 a cobertura florestal no mundo era de 4 bilhões de hectares, o que corresponde a 31% de toda a superfície do planeta Terra, sendo o Brasil o segundo maior país com cobertura florestal absoluta no mundo (somando florestas plantadas e naturais), com aproximadamente 519 milhões de hectares, ficando atrás apenas da Rússia que apresenta 809 milhões de hectares. Essas florestas cobrem cerca de 62% do território brasileiro e 49% do território russo. Os outros três países que completam o grupo de países com maior cobertura florestal são Canadá com 310 milhões de hectares, Estados Unidos com 304 milhões de hectares e China com cerca de 207 milhões de hectares. O Brasil possui a sétima maior área com florestas plantadas e é o terceiro maior produtor mundial de madeira destinada ao setor energético (FAO, 2011).

Digitized by Google

Apesar da vocação florestal, o Brasil ainda não explora todo o seu potencial produtivo, tanto para o cultivo quanto para a geração de produtos derivados de madeira (LONGUE & CODOLLETE, 2013). Como mostrado na Tabela 6, apesar de ser um país de dimensões continentais, o Brasil possui proporcionalmente menos áreas de florestas plantadas que outros pequenos países como a Finlândia e o Japão e apresenta uma área total de plantio de florestas muito menor que os Estados Unidos e a China, seus tradicionais competidores. Na América do Sul, além do Brasil, apenas o Chile possui florestas plantadas. A Tabela 7 apresenta detalhadamente os dados mais recentes de florestas plantadas no mundo.

A produção mundial de madeira bruta, entre 1989 e 2009 foi superior a 130 milhões de metros m^3, sendo mais da metade destinada para biomassa (energia) 58%, em seguida a indústria de serraria e laminação que ocupou a segunda posição 31%. Os outros 11% da madeira produzida no mundo foram distribuídos entre as indústrias de celulose e outros segmentos (MOREIRA, 2011).

Em alguns países o consumo de lenha como fonte de energia pode ser utilizado como indicador do nível de subdesenvolvimento econômico da região (BRITO, 2007; VIDAL; HORA, 2011), devido a forte associação negativa entre a participação da biomassa sólida na matriz energética e o PIB *per capita*, isto é, observou-se durante o período do estudo que quanto maior o nível de desenvolvimento do país menor a participação da lenha na matriz energética. Entretanto, fatores como a pressão de organismos governamentais por fontes menos prejudiciais a saúde do Planeta e a necessidade de sustentar as demandas futuras de energia, tem destacado a madeira como uma alternativa viável de energia, revertendo à associação com o subdesenvolvimento.

Digitized by Google

Tabela 7 - Comparativo entre áreas de florestas plantadas no mundo

Países	Extensão do território (1000 ha)	Área total Floresta* (1000 ha)	Área floresta plantada (1000 ha)	Área floresta plantada/ Área total floresta	Área floresta plantada/ Território
China	942.530	206.860	77.157	37,3%	8,2%
EUA	916.193	304.022	25.363	8,3%	2,8%
Rússia	1.638.139	809.091	16.991	2,1%	1,0%
Japão	36.450	24.979	10.326	41,3%	28,3%
Índia	297.319	68.434	10.211	14,9%	3,4%
Canadá	909.351	310.134	8.963	2,9%	1,0%
Brasil	832.512	519.523	7.418	1,4%	0,9%
Finlândia	30.409	22.156	5.904	26,6%	19,4%
Alemanha	34.877	11.076	5.283	47,7%	15,1%
Suécia	41.033	28.203	3.613	12,8%	8,8%
Indonésia	181.157	94.432	3.549	3,8%	2,0%
Vietnã	31.008	13.797	3.512	25,5%	11,3%
Turquia	76.963	11.334	3.418	30,2%	4,4%
México	194.395	64.802	3.203	4,9%	1,6%
Espanha	49.919	18.173	2.680	14,7%	5,4%
Chile	74.880	16.231	2.384	14,7%	3,2%

*(Primaria + Regenerada + Plantada)
Fonte: (1) FAO (2011); (2) ABRAF (2012)

A utilização de *pellets* de madeira é um exemplo de ampliação do uso da biomassa florestal na matriz energética de países desenvolvidos, principalmente devido às demandas recentes criadas por políticas e metas de uso de bioenergia na Europa. Em 2012 a produção mundial foi de 19 milhões de toneladas, com cerca de metade (9,3 milhões de toneladas) comercializado internacionalmente. A Europa e América do Norte são responsáveis por quase toda a produção mundial de *pallets* (66% e 31%, respectivamente) e consumo (80% e 17%, respectivamente), segundo dados da FAO, 2012.

3.2 Panorama do setor florestal no Brasil

De acordo com MONTEBELLO & BACHA (2009), historicamente o processo de evolução da silvicultura nacional pode ser dividido em três etapas. A primeira correspondeu ao período que

Digitized by Google

vai do descobrimento do Brasil até o início dos incentivos fiscais ao reflorestamento em 1965, a segunda fase considera o período de vigência dos incentivos fiscais de 1966 a 1988 e a terceira fase corresponde do período pós-incentivos fiscais a reflorestamentos que vai de 1988 até hoje.

O estabelecimento do programa de incentivos fiscais (FISET) do governo federal, instituído na década de 1960 no âmbito do Plano Nacional de Desenvolvimento, viabilizou a execução de plantios florestais de larga escala no Brasil. Tais incentivos foram criados através das Leis 5.106, de 09/1966, e 1.134, de 11/1970, visando expandir o setor florestal a fim de se garantir matéria-prima para viabilizar as indústrias de papel e celulose e siderurgias (MENDES, 2005). Esse programa concedia incentivos via o abatimento de alíquotas de imposto em até 50% para os produtores que comprovassem a aplicação dos recursos em reflorestamentos. O FISET também contribuiu com investimentos e ampliação da infraestrutura tecnológica do setor (JUVENAL; MATTOS, 2002; COUTO; MULLER, 2010). Contudo, o FISET foi extinto em 1988 (MENDES, 2005). Além dos incentivos do FISET, a promulgação do primeiro Código Florestal em 1965 e a criação do Instituto Brasileiro de Desenvolvimento Florestal (IBDF) em 1967 foram ações que ajudaram a definir uma nova política florestal para o País, promovendo o reflorestamento em larga escala (MENDES, 2005).

O período que compreende o fim da primeira fase e o início da segunda fase (meados da década de 80) destacou-se pelo surgimento dos primeiros centros de pesquisa florestais e as inovações tecnológicas nas áreas de melhoramento genético, introdução de novas espécies e processamento industrial da madeira considerados pontos chave para o se alcançar ganhos de produtividade (MONTEBELLO & BACHA, 2009).

Em 2002, o governo federal lançou o Programa de Plantio Comercial de Florestas (PROPFLORA), coordenado pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA) no âmbito do Plano Agrícola 2002/200 para apoiar pequenos e médios produ-

Digitized by Google

tores rurais a implantarem e manterem suas florestas comerciais, com recursos técnicos e mão-de-obra adequados. Para os produtores familiares, o MDA (Ministério do Desenvolvimento Agrário) estendeu os financiamentos do Programa Nacional de Agricultura Familiar para as atividades florestais, criando em julho de 2002 o PRONAF FLORESTAL (MENDES, 2005). Outras ações governamentais, com maior ou menor impacto sobre o desenvolvimento e o manejo florestal sustentável, incluem o Plano Plurianual do Governo Federal (PPA 2004-2007), o Programa Nacional do Meio Ambiente (PNMA), o Programa Piloto para Proteção de Florestas Tropicais no Brasil (PPG7) e o Programa Nacional da Diversidade Biológica (PRONABIO).

Atualmente uma nova rodada de incentivos aos reflorestamentos está em andamento com ações como o Plano Setorial de Mitigação e Adaptação às Mudanças Climáticas para a Consolidação de uma Economia de Baixo Carbono (Plano ABC) que prevê o plantio de três milhões de hectares de florestas plantadas até 2020. Dentre as espécies florestais cultivadas com a finalidade industrial no Brasil, o eucalipto é a que mais se destaca com 70,8% da área plantada, sendo a produtividade brasileira considerada a maior do mundo (STCP, 2012). Em segundo lugar aparece o gênero pinus com 22% da área plantada. Os outros 7,2% das áreas plantadas são ocupados por outras espécies (Acácia, Seringueira, Paricá, Araucária, Téca, Pópulus, etc.). Segundo dados da Associação Brasileira de Produtores de Florestas Plantadas (ABRAF, 2013), os estados brasileiros com maior área absoluta de florestas plantadas em 2012 foram Minas Gerais, São Paulo, Paraná e Bahia, que juntos representam 61,7% da área de floresta plantada do país. A região Sudeste se destaca pela maior concentração dos plantios florestais (53%), seguida pela região Sul (11,5%), Centro Oeste (13,4%), Nordeste (15,8%) e Norte (6,2%) (ABRAF, 2013). Para o melhor entendimento da composição das áreas plantadas por estado, a Tabela 8 apresenta a evolução da área plantada com reflorestamentos de pinus e eucalipto.

Digitized by Google

Tabela 8 - Evolução da área de florestas plantadas por estado (Eucaliptos e Pinus)

UF	2006	2007	2008	2009	2010	2011	2012	Taxa
MG	1.327.429	1.361.607	1.423.210	1.440.000	1.536.310	1.477.195	1.491.681	1,0%
SP	1.130.332	1.121.529	1.173.560	1.197.330	1.206.818	1.188.403	1.186.497	-0,2%
PR	808.361	824.648	857.320	853.710	847.931	846.860	817.566	-3,6%
BA	594.992	591.348	622.700	659.480	658.034	628.960	616.694	-2,0%
SC	601.333	622.045	628.660	650.990	647.992	642.941	645.965	0,5%
RS	365.623	404.623	450.480	443.190	441.997	445.004	449.533	1,0%
MS	147.819	228.384	284.050	307.760	392.042	487.399	597.135	18,4%
ES	212.208	212.912	214.400	208.510	207.431	200.058	205.895	2,8%
PA	115.955	126.387	136.300	139.720	148.656	151.378	159.657	5,2%
MA	93.285	106.802	111.120	137.360	151.403	165.717	173.324	4,4%
GO	64.046	65.107	72.080	73.140	70.679	70.384	54.513	-29,1%
AP	78.963	67.874	64.930	63.690	49.384	50.543	49.951	-1,2%
MT	46.153	57.158	58.590	61.540	61.950	58.843	59.980	1,9%
TO	13.901	22.355	32.770	45.160	48.392	66.352	109.853	39,6%
PI	-	-	-	-	-	-	27.730	-
RJ	-	-	-	-	-	-	18.368	-
Outros							470	-

Fonte: ABRAF (2013)

Observa-se que os estados que apresentaram maior crescimento em relação ao ano de 2011 foram Mato Grosso do Sul (18,4%) e Tocantins (39,6%). Entre os estados que apresentaram redução em relação ao ano de 2011, os estados da Bahia, Paraná e Goiás apresentaram os maiores índices. Atualmente aproximadamente 61% do território é composto por áreas preservadas e aproximadamente 0,9% por áreas com florestas plantadas conforme a Tabela 9, sendo os 38% restantes compostos por área utilizadas para a pecuária, agricultura para produção de alimentos e áreas urbanas.

Tabela 9 - Usos da terra no Brasil

Usos do solo	Área em (mil Ha)	%
Florestas nativas em áreas não privadas	423,3	49,7
Florestas nativas em propriedades rurais	93,9	11,0
Florestas Plantadas	6,9	0,9
Pecuária	160	18,8
Agricultura / Produção de alimentos	71	8,3
Cidades e Infraestrutura	95,8	11,3

Fonte: Adaptado de BRACELPA (2013)

Digitized by Google

A distribuição das áreas de plantio conforme as espécies e usos específicos são mencionados na Tabela 10.

Tabela 10 - Composição das florestas plantadas no Brasil em 2012 (UF)

Espécie	Principais Usos	Principais Estados	Área (Ha)	%
Eucalipto (Eucalyptus spp*)	Madeira: Energia, carvão, cavaco p/ celulose, painéis de madeira, dormentes, postes, construção civil, óleos essenciais.	MG, SP, BA, ES, MS, RS, PR, SC, PA e MA.	5.102.030	71
Pinus spp	Madeira: energia, carvão, cavaco p/ celulose, painéis de madeira, forros, ripas, móveis. Resina: tintas, vernizes, solventes.	PR, SC, RS, SP e MG.	1.562.782	21,75
Acácia (Acacia mearnsii e Acacia mangium)	Madeira: energia, carvão, cavaco p/ celulose, painéis de madeira. Tanino: curtumes, adesivos, petrolífero, borrachas.	RS e RR	148.311	2,12
Seringueira (Hevea brasiliensis)	Madeira: energia, celulose. Seiva: borracha	AM	168.848	2,36
Paricá (Schizolobium amazonicum)	Lâmina e compensado, forros, palitos, papel, móveis, acabamentos e molduras.	PA e MA	87.901	1,22
Teca (Tectona grandis)	Construção civil (portas, janelas, lambris, painéis, forros), assoalhos e decks, móveis, embarcações e lâminas decorativas.	MT, AM, AC.	67.329	0,97
Araucaria angustifolia	Serrados, lâminas, forros, molduras, ripas, caixotaria, estrutura de móveis, fósforo, lápis e carretéis.	PR e SC	11.343	0,16
Populus spp.	Fósforos, partes de móveis, portas, marcenaria interior, brinquedos, utensílios de cozinha	PR e SC	4.216	0,06
Outras			33.183	0,46
Total			**7.185.943**	**100**

Fonte: SFB (2013) e ABRAF (2013)

3.3 Florestas plantadas e energia

As atividades florestais podem causar impactos positivos e negativos ao meio ambiente dependendo da forma como são utilizados os recursos naturais envolvidos na implantação do empreendimento florestal. A busca por fontes alternativas aos combustíveis fósseis tem incentivado a formação de um novo modelo de produtividade florestal com grande volume de árvores por hectare e menor tempo de rotação. A demanda por madeira no Brasil vem aumentando, particularmente a demanda por biomassa para a produção de energia e deve continuar a aumentar segundo pesquisas do setor florestal (SFB, 2013), sendo que pesquisas sobre sustentabilidade das plantações florestais são quase inexistentes

Digitized by Google

FOELKEL (2012). No setor de produção de energia, a madeira, que também, pode tomar a forma conhecida por lenha, sempre teve grande importância para o desenvolvimento da humanidade por seu custo baixo e fácil acesso, sendo utilizada na preparação de alimentos. Ao longo dos tempos, passou a ser utilizada como combustível em processos para geração de energia térmica, mecânica e elétrica (BRITO, 2007). No Brasil a madeira consumida para energia tem sido historicamente relacionada à produção de carvão vegetal, aos consumos residencial, industrial e agropecuário (Idem, 2007).

Segundo dados da FAO (2011), os cinco principais países produtores de madeira para energia entre os anos 1989 e 2009 foram: Índia (6,08 bilhões de m^3), China (4,95 bilhões de m^3), Brasil (2,75 bilhões de m^3), Indonésia (1,93 bilhões de m^3) e Etiópia (1,51 bilhões de m^3); sendo responsáveis por 45,14% da produção mundial. Pode-se observar que os grandes produtores mundiais de madeira para energia são países em desenvolvimento. Ocorre um aumento tanto na produção, quanto no consumo de madeira para energia por parte destes países, particularmente a Etiópia, o Brasil e a Índia que apresentaram taxa média de crescimento anual de 1,64%, 0,87% e 0,59% respectivamente.

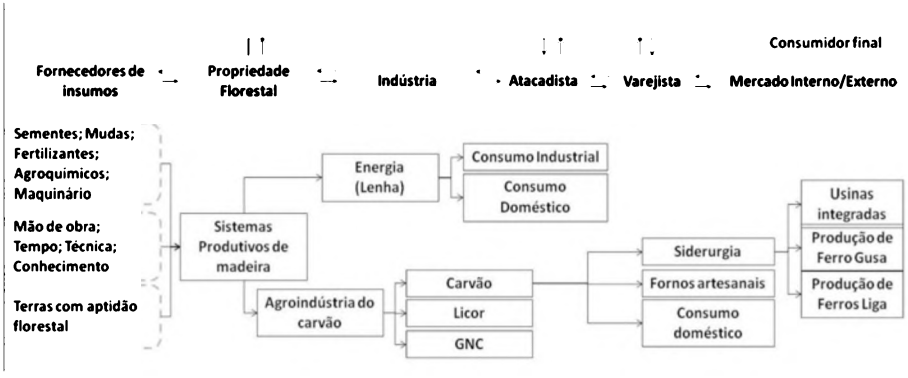
No Brasil a produção de lenha diminuiu cerca de 7% no período de 2005 a 2011, segundo dados do (IEA, 2013), reflexo principalmente da queda na utilização de lenha pelo setor residencial, ao contrário do setor industrial que apresentou aumento no consumo devido ao aumento no número de empresas que utilizam madeira para a geração de energia elétrica. Na distribuição desta madeira, 1% foi utilizada para o setor madeireiro gerar sua própria energia, 37% para a produção de carvão vegetal e 62% foram utilizados para fins energéticos em outros setores como econômicos, como o setor industrial que utilizou 45%. Na distribuição da utilização de lenha no setor industrial destacam-se os seguimentos de cerâmica (32%), alimentos e bebidas (31%) e de papel e celulose (21%). A produção de carvão se destaca, em decorrência da de-

Digitized by Google

manda existente pelo produto junto ao setor siderúrgico. O Brasil é o maior produtor mundial de aço produzido com o emprego do carvão vegetal para fins de redução do minério de ferro. Trata-se de um setor consolidado e em expansão, gerador de centenas de milhares de postos de trabalho, grande quantidade de impostos e de renda (Brito, 2007).

A lenha *in natura* e o carvão vegetal são os principais produtos da cadeia produtiva agroindustrial da madeira para energia, todavia, pesquisas sobre novos produtos energéticos estão sendo intensificadas. A produção de bio-óleo, celulignina, gases de síntase e álcool vêm gradativamente ganhando importância, assumindo papéis estratégicos e papeis complementares aos usos consagrados de uso da madeira para fins energéticos (FONTES, 2005; SANTOS, 2012). Também, estão em andamento pesquisas sobre geração de hidrogênio a partir de biomassa, geração de etanol por rota biotecnológica com utilização de resíduos florestais (EMBRAPA, 2013). Pode-se verificar na Figura 8 o esquema da cadeia produtiva agroindustrial da madeira para energia.

Figura 8 - Representação esquemática da cadeia produtiva de madeira para energia



Fonte: Adaptado de FONTES (2005) e ABRAF (2013)

O elo dos fornecedores de insumos está em evolução, com desenvolvimento de novos genótipos, disponibilização de sementes e mudas, fertilizantes e máquinas de qualidade, tendo como gargalo o alto custo.

Digitized by Google

A madeira para conversão energética pode seguir diversos canais: podem ser destinadas ao consumidor final, aos atacadistas, aos varejistas, ou aos centros de transformação (carvoarias, termelétricas, etc.) sendo transformada em uma fonte secundária de energia (carvão e eletricidade). Segundo dados da ANEEL (2014), das 3.144 usinas de energia elétrica que existem no país, 8% utilizam biomassa e, no período de 2005 a 2014, esse segmento cresceu 227%. A participação da madeira entre as biomassas utilizadas, embora tenha uma participação pouco expressiva (3,6%) teve o número de usinas mais que duplicado no período de 2005 e 2014, passando de 23 para 51 respectivamente. Isso permitiu um aumento de 89% da capacidade instalada (Tabela 11).

Tabela 11 - Composição da matriz elétrica nacional - por fonte: 2005 a 2014

Fontes de Energia	2005		2011		2014		
	N de usinas	Capacidade instalada (kW)	N de usinas	Capacidade instalada (kW)	N de usinas	Capacidade instalada (kW)	%
Não renovável (subtotal)	578	18.333.307	1.197	26.076.374	1.326	27.273.315	20
Gás	95	10.080.701	148	13.684.493	157	14.281.944	10,47
Petróleo	474	4.830.606	1.035	7.720.553	1.154	7.611.906	5,58
Nuclear	2	2.007.000	2	2.007.000	2	1.990.000	1,46
Carvão Mineral	7	1.415.000	12	2.664.328	13	3.389.465	2,48
Renovável (subtotal)	1.101	77.261.455	1.958	105.228.182	1.816	101.007.568	74
Hidro	587	70.977.169	1.050	84.690.402	1.106	86.698.054	63,54
Biomassa (subtotal)	257	3.142.193	454	10.268.890	478	11.423.613	8,37
Bagaço de Cana	218	2.249.259	367	8.532.612	378	9.339.426	6,84
Licor Negro	12	665.572	14	1.246.222	16	1.530.182	1,12
Madeira	23	200.832	45	379.235	51	432.635	0,32
Biogás	2	20.030	19	74.388	24	84.937	0,06
Casca de Arroz	2	6.400	9	36.433	9	36.433	0,03
Eólica	-	-	-	-	135	2.876.576	2,11
Fotovoltaica	-	-	-	-	97	9.325	0
Importação	-	-	-	-		8.170.000	7,89
Total (Brasil)		**92.452.669**	**2.701**	**121.035.666**	**3.144**	**136.455.233**	**100**

Fonte: Agência Nacional de Energia Elétrica – ANEEL
Dados de abril de 2014. Disponível em: <http://www.aneel.gov.br>

Ressalta-se que dos polos de produção da madeira utilizadas nas usinas, Santa Catarina, Paraná e São Paulo utilizam florestas

Digitized by Google

plantadas para suprir a demanda, enquanto os polos localizados no Pará, Mato Grosso e Rondônia utilizaram madeira oriunda de florestas nativas e plantadas.

Outro aspecto da cadeia de madeira para energia são as atividades de apoio presentes em todos os segmentos da cadeia: Sistema Financeiro, Políticas governamentais, Logística, ONGs, Extensão Rural e Associações, Sindicatos e outras (FONTES, 2005).

3.1.1 Florestas energéticas no Tocantins

O estado do Tocantins localiza-se no centro geográfico do País, possibilitando fazer limites com estados do Nordeste, Centro-Oeste e do próprio Norte. Sua vegetação é composta principalmente pelo bioma Cerrado, que corresponde por cerca de 90% da superfície do Estado e pelo bioma Amazônia que localizado em uma área de transição na divisa com o estado do Pará, corresponde a cerca de 9% da superfície do estado.

Segundo dados da Secretaria de Meio Ambiente e Desenvolvimento Sustentável do Estado do Tocantins (SEMADES-TO), em 2007 a cobertura vegetal remanescente era de 66,5% da área do estado, conforme a Tabela 12, que apresenta de forma resumida uma comparação entre a cobertura vegetal natural original e remanescente em 2007. Nota-se também grande perda da área ocupada por formações de Floresta Ombrófila (83%), que juntamente com as formações de Floresta Ombrófila/Estacional e Floresta Estacional são as áreas de maior aptidão para o manejo florestal voltado a produção agroindustrial segundo (SEMADES--TO, 2013). Estas formações fitoecológicas representam cerca de 15% da superfície do estado ou aproximadamente 4 milhões de hectares.

Digitized by Google

Tabela 12 - Comparativo da Cobertura Vegetal Natural Original e Remanescente - Tocantins (2007)

Tipo Principal	Original		Remanescente (2007)		Porcentagem de alteração até 2007
	Área (1.000 ha)	%	Área (1.000 ha)	%	
Floresta Ombrófila	2.058,0	7,4	354,7	1,3	83%
Floresta Ombrófila-estacional	496,0	1,8	219,8	0,8	56%
Floresta estácional	4.576,1	16,5	3.550,7	12,8	22%
Formações Pioneiras	80,9	0,3	72,1	0,3	11%
Savana/Cerrado	16.886,8	60,8	12.156,6	43,8	28%
Savana/Floresta –estacional	3.205,1	11,5	2.100,0	7,6	34%
Praias/Dunas	18,4	0,1	18,4	0,1	0%
Subtotal	**27. 321,3**	**98,4**	**18.475,3**	**66,5**	**32%**
Outras coberturas e Ocupações	440,8	1,6	9.286,8	33,5	-
Total Geral	**27.762,1**	**100,0**	**27.762,1**	**100,0**	0%

Fonte: SEMADES-TO, 2013

O Estado do Tocantins contava em 2011 com uma área total de plantios florestais de 83,6 mil hectares, incluindo plantios para fins madeireiros e não madeireiros. Como se observa na Tabela 13, a área total com florestas plantadas cobria cerca de 0,3% da superfície do estado, a área plantada para fins madeireiros corresponde a quase 81 mil hectares, sendo o eucalipto espécie mais plantada (76 mil ha).

Considerando o ritmo de crescimento dos plantios florestais no estado, e estudo diagnóstico do setor florestal realizado pela SEAGRO-TO, em parceria com a UFT, estima-se que a área plantada total em 2014 seja de aproximadamente 155 mil hectares.

Digitized by Google

Tabela 12 – Áreas com Plantios Florestais, por Grupo de Espécie, no Estado do Tocantins em 2011

Destinação	Gênero	Área plantada (ha)	% dos Plantios	% do Estado
	Eucalipto	75.977,8	90,9	0,27
	Teca	2.798,0	3,3	0,01
	Pinus	850,0	1,0	0,00
Madeireiros	Acácia	650,2	0,8	0,00
	Guanandi	400,0	0,5	0,00
	Mógno Africano	89,2	0,1	0,00
	Paricá	2,5	0,0	0,00
	Subtotal	**80.767,7**	**96,6**	**0,29**
	Seringueira	1.840,0	2,2	0,01
	Pupunha	255,0	0,3	0,00
Não madeireiros	Neem	188,0	0,2	0,00
	Pequi	9,5	0,0	0,00
	Outros	530,0	0,6	0,00
	Subtotal	**2.822,5**	**3,4**	**0,01**
	Totais Plantações	**83.590,2**	**100,0**	**0,30**
	Total Geral Estado	**27.767.100**		**100,0**

Fonte: SEAGRO, 2012; STCP, 2013

Entre as espécies com fins madeireiros destinados a produção de energia, que são o foco deste trabalho, o eucalipto é a que mais se destaca, seguido pelas espécies Pinus e Acácia. Cerca de 50% dos plantios com eucalipto concentra-se em 5 municípios principais. A distribuição dos plantios por espécie e principais municípios onde as mesmas estão plantadas é apresentada na Tabela 14.

A região Norte do estado do Tocantins conhecida como Bico do Papagaio possui a maior concentração de plantios. Os municípios de Araguatins e São Bento do Tocantins concentram grande parte destes plantios, com ciclos de curta rotação, isto é, cortes previstos aos 7 anos de idade com foco no processamento industrial, geração e energia e carvão vegetal. A estimativa de produtividade dos plantios é de 30 m^3/ha/ano, com a expectativa de aumento à medida que pesquisas no melhoramento genético da cultura sejam realizados. Entre as espécies madeireiras plan-

Digitized by Google

tadas no estado do Tocantins estão a Teca, o Guanandi, o Mogno Africano e Paricá. A Teca, apesar de não ter como especialidade a produção de energia, representa a segunda espécie florestal mais cultivada no estado, com quase 2,8 mil hectares, o que corresponde a 3,4% da área com plantações florestais do Estado.

Tabela 14 - Distribuição dos Plantios para fins Madeireiros e energéticos1 - Tocantins (2011)

	Município	Área plantada (ha)	%
Eucalipto	Araguatins	13.145,0	15,7
	São Bento	7.100,00	8,5
	Ananás	6.525,4	7,8
	Wanderlândia	5.051,0	6,0
	Darcinópolis	2.511,7	3,0
	Outros	41.644,9	49,8
Subtotal Eucalipto		**75.978,0**	**90,9**
Pinus	Almas	450,0	0,54
	Porto Nacional	400,0	0,48
Subtotal Pinos		**850,0**	**1,02**
Acácia	Natividade	265,9	0,32
	Juarina	194,0	0,23
	Paranã	110,4	0,13
	Conceição do Tocantins	80,0	0,10
Subtotal Acácia		**650,3**	**0,78**
Total Demais espécies		**1.992,0**	**2,38**
Total Plantações		**83.590,2**	**100,0**

[1] Não incluem espécies apenas madeireiras e não madeireiras
Fonte: Adaptado de SEMADES-TO (2013)

Já entre as espécies não madeireiras, isto é, plantios destinados produção de extrativos, como por exemplo, a borracha, fibras, gomas não elásticas entre outros, destaca-se a Seringueira, representando 80% dos plantios não madeireiros no Estado. A expansão da área de cultivo é resultado de incentivos aos pequenos produtores rurais por meio do "Programa de Estruturação de Polos de Produção de Seringueira", promovido pelo governo estadual. Entre as demais espécies não madeireiras cultivadas no Estado estão: Pupunha, Neem, Pequi e Babaçú.

Digitized by Google

A Figura 9 apresenta o mapa com as principais espécies plantadas e regiões com maior potencial de manejo sustentável. O mapa também apresenta a localização de alguns dos mercados consumidores regionais atendidos pela produção florestal do Tocantins.

Como visto anteriormente, entre as espécies de plantios utilizadas também com a finalidade energética o Eucalipto se destaca entre as demais, podendo-se agrupar sua produção em dois polos principais no Estado:

- **Polo da região Sul:** concentra nos municípios de Peixe, Dueré, Brejinho do Nazaré, São Valério, Aliança e Natividade. Concentra 30% da produção estadual sendo a maior parte destinada a produção de celulose.
- **Polo da região Norte (Bico do Papagaio):** abrange desde a divisa com o Maranhão, na região de São Miguel do Tocantins, até Wanderlândia e Araguaína. A região concentra cerca de 63% da produção estadual de eucalipto. A produção é destinada principalmente para a produção de celulose e uso na siderurgia.

A produção de carvão vegetal no Tocantins é comparativamente pequena e bastante instável. Praticamente toda a produção de carvão vegetal do Estado do Tocantins é proveniente de floresta nativa e consumida em outros estados.

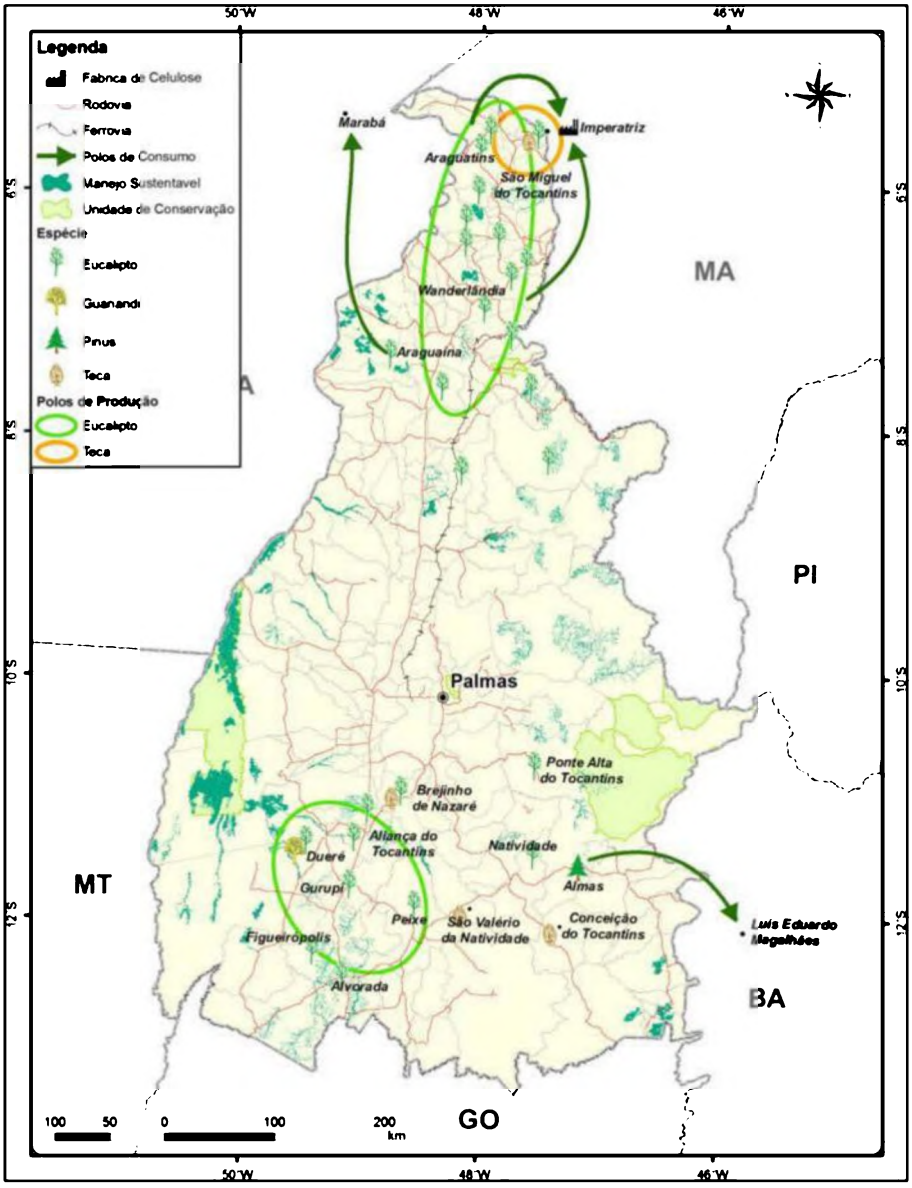
No Tocantins existem unidades produtoras de carvão para fins industriais voltadas a atender a demanda da indústria siderúrgica, principalmente no Polo siderúrgico do Maranhão.

Para atender uma demanda cada vez mais crescente e devido a pressão sobre o uso de carvão a partir de fonte nativa, empresas com sede no Maranhão e Pará mantêm plantios cada vez maiores nos estados do Pará e Maranhão, visando à produção futura de carvão vegetal no médio e longo prazo.

Outro segmento de utilização da madeira para a energia e que tem apresentado crescimento é a secagem de grãos. Esse aumento pode ser explicado pelo crescimento das áreas agrícolas na região norte do estado, com destaque para a região da MAPITOBA (Bahia, Maranhão, Piauí e Tocantins).

Digitized by Google

Figura 9 - Principais Polos de Produção e Consumo de Florestas Plantadas no Tocantins



Fonte: SEMADES-TO (2013)

3.4 Plano ABC

A crescente preocupação com as mudanças do clima, em par-

Digitized by Google

ticular com o "Aquecimento Global" colocam o Brasil em posição de destaque nas discussões sobre o tema, principalmente depois que o país assumiu compromissos voluntários de redução de Gases do Efeito Estufa entre 36,1% e 38,9% até 2020 (BRASIL, 2012).

O Brasil tem a sua disposição diversas tecnologias sustentáveis de baixa emissão de carbono, desenvolvidas para as diferentes condições tropicais e subtropicais, principalmente, para a agropecuária, como pode ser obervado na Tabela14.

O compromisso de redução das emissões de gases do efeito estufa (GEE) foi estabelecido durante a 15ª Conferencia das Partes (COP 15), realizada em 2009, na cidade de Copenhagen, na Dinamarca, onde se estimou reduções da ordem de um bilhão de toneladas de CO_2 equivalente (t CO_2 eq). No caso específico da agricultura o potencial estimado para redução é em torno de 133 a 162 milhões de $MgCO_2eq$ (BRASIL, 2012).

O Plano Setorial de Mitigação e de Adaptação às Mudanças Climáticas para a Consolidação de uma Economia de Baixa Emissão de Carbono na Agricultura - Plano ABC é um dos planos setoriais abordados no artigo 3º do Decreto nº 7.390/2010.

A finalidade do plano é o detalhamento das ações para a adoção de tecnologias sustentáveis que auxiliem o país no cumprimento do compromisso de redução de emissão de gases do efeito estufa assumido.

As principais ações do Plano ABC são apresentadas na Tabela 15. Foram divididas em 7 programas, sendo 6 referentes a processos tecnológicos de mitigação e 1 referente a ações de adaptação às mudanças do clima.

Em cada um dos sete programas foram adotadas ações de apoio como, por exemplo, o mapeamento de áreas com pastagens degradadas, zoneamento de pastagens na Amazônia legal, mapeamento de áreas com potencial para implantação de iLPF, mapeamento de áreas com potencial para implantação de SAFs, mapeamento de áreas estratégicas para a implantação de SPD e mapeamento de áreas potenciais para a implantação de florestas plantadas.

Digitized by Google

O Brasil possui cerca de 50 milhões de hectares com pastagens degradadas, onde o pasto não é cultivado e o gado é levado para outro lugar quando acaba a grama natural. Em termos comparativos, a recuperação de 19 milhões de hectares de pastagens degradadas diminuiria a emissão de carbono em 132 milhões de tCO_2eq até o fim da década, isto é, 13% do total de reduções prometidos na COP-15.

Tabela 15 - Processos Tecnológicos, compromisso nacional relativo e potencial de mitigação por redução de emissão de GEE

Processo Tecnológico	Compromisso (aumento de área/uso)	Potencial de Mitigação (milhões Mg CO_2 eq)
Recuperação de Pastagens Degradadas[1]	15,0 milhões ha	83 a 104
Integração Lavoura – Pecuária – Florestas[2] (iLPF)	4,0 milhões ha	18 a 22
Sistema de Plantio Direto (SPD)	8,0 milhões ha	16 a 20
Fixação Biológica de Nitrogênio	5,5 milhões ha	10
Florestas Plantadas[3]	3,0 milhões ha	-
Tratamento de dejetos animais	4,0 milhões m^3	6,9
Total		**133 a 162,9**

1 Por meio de manejo adequado e adubação
2 Incluindo os sistemas agroflorestais (SAFs)
3 Não está computado o compromisso brasileiro relativo ao setor de siderurgia; e não foi contabilizado o potencial de mitigação de emissão de GEE
Fonte: BRASIL (2012)

Segundo Dias Filho (2011) a degradação de pastagens é um problema de escala global e que no Brasil tem sido assunto de discussão por causa dos importantes prejuízos econômicos e ambientais.

O processo de degradação de pastagem é definido por Dias Filho (2011), como um fenômeno complexo que leva a gradativa diminuição da capacidade de suporte da pastagem de sustentar os níveis de produção exigidos pelos animais, assim como, de superar os efeitos de pragas e doenças invasoras acarretando na degradação dos recursos naturais.

Entre as principais causas estão: as falhas no uso do pasto como, práticas de manejo inadequadas, a ausência de adubação, uso demasiado do fogo, fatores bióticos como ataques de insetos-

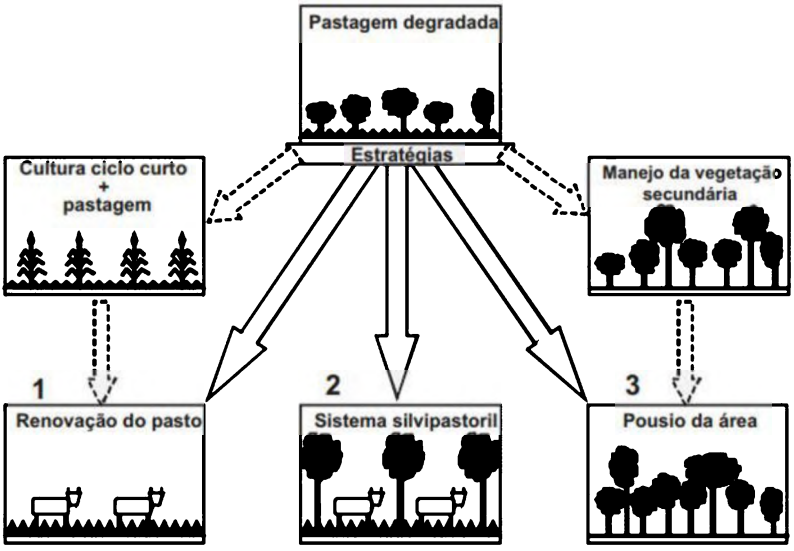
Digitized by Google

-praga ou doenças (patógenos) e fatores abióticos como excesso ou falta de chuva, baixa fertilidade de drenagem deficiente do solo (DIAS FILHO, 2011).

O impacto de cada uma dessas causas na degradação das pastagens brasileiras varia de acordo com a região geográfica, assim como, variam também as estratégias de renovação da pastagem que segundo Dias Filho (2011), podem ser classificada em três linhas principais, conforme a Figura 10: Renovação (reforma) da pastagem; Implantação de sistemas agrícolas e agroflorestais e Pousio da pastagem.

No caso da renovação de pastagem, quando a cobertura do solo pelas plantas forrageiras é muito baixa (<50%), ações com o objetivo de restabelecimento o pasto podem ser aplicadas. Estas ações estão condicionadas a fatores como o tamanho (pequenas ou grandes propriedades) da área a ser recuperada, as espécies predominantes, o estágio de desenvolvimento de plantas daninhas e o capital disponível para renovação.

Figura 10 - Estratégias para a recuperação de produtividade de pastagens degradadas



Fonte: DIAS FILHO (2011)

Digitized by Google

Dias Filho (2011) também sugere um sistema agrícola e um sistema agroflorestal como alternativa para a recuperação de pastagens degradas. Sistemas agrosilvipastoris: São sistemas que associam árvores, campos de cultivo de culturas anuais em sistemas de rotação ou de consórcio e animais. Ele consegue ao mesmo tempo conservar os recursos naturais, aumentar a produtividade agrícola e pecuária, fixar o homem no campo, trazendo melhoria na qualidade de vida (ORTEGA *et al.*, 2008).

Embora a adoção desse sistema implique em investimentos em tecnologia por parte dos produtores provenientes da aquisição de máquinas, implementos para plantio e infraestrutura de armazenamento de grãos, a integração dos sistemas também tem como pontos positivos a elevação dos níveis de produtividade e diversidade da propriedade rural e consequentemente a diminuição dos desmatamentos.

Sistemas silvipastoris (plantados ou com manejo da vegetação nativa secundária): são a combinação intencional de árvores, pastagem e gado numa mesma área ao mesmo tempo e manejados de forma integrada. O plantio das árvores pode ter fins produtivos ou pode-se adotar o processo de regeneração natural de espécies arbóreas, promovendo o aumento da diversidade biológica e a conservação dos nutrientes e da água nas áreas produtivas.

O Pousio da pastagem refere-se à estratégia de recompor a área de reserva legal da propriedade ou recuperar áreas que não deveriam ter sido originalmente desmatadas (tal como margens de cursos d'agua e áreas com drenagem deficiente) por meio de processo natural de sucessão secundária ou por meio do controle seletivo da vegetação com o objetivo de dificultar o desenvolvimento de certas plantas indesejáveis e facilitar o estabelecimento daquelas consideradas desejáveis como, por exemplo, árvores com maior capacidade de crescimento e de acúmulo de nutrientes na biomassa, ou ainda espécies de maior valor econômico.

Segundo Castro *et al.*, (2006) *apud* Dias Filho, (2011) quando comparado com outras modalidades de uso e ocupação do solo

Digitized by Google

o reflorestamento ou plantio de árvores é a atividade mais recomendada para a recuperação de áreas degradas, conservação do solo e recomposição da paisagem. As áreas de acentuada declividade devem ser protegidas com uma cobertura vegetal e merecem uma atenção especial quanto aos riscos de erosão e o arraste do solo.

Considerando a grande extensão de áreas de pastagens degradadas no Brasil e a necessidade de métodos analíticos de custo reduzido, o projeto GeoDEGRADE[1] coordenado pela Embrapa Monitoramento por Satélite (Campinas-SP), foi elaborado visando desenvolver instrumentos baseados em tecnologias geoespaciais que permitam a identificação das áreas com pastagens degradadas, isto é, áreas com potencial para aumento da produção agropecuária. Além da Embrapa Monitoramento por Satélite, o projeto tem participação de outras seis unidades da Embrapa, da Agência Paulista de Tecnologia dos Agronegócios (APTA), do Centro de Cooperação Internacional em Pesquisa Agronômica para o Desenvolvimento (CIRAD), da França, do Instituto Nacional de Pesquisas Especiais (INPE), da Unicamp e do Museu Paraense Emílio Goeldi (EMBRAPA, 2012).

Entre 2011 e 2012 foram realizadas duas campanhas, ambas no âmbito do Rally da Pecuária[2], expedição de cunho exploratório organizada pelas consultorias Agroconsult e Bigma, a fim de diagnosticar a condição das pastagens nos biomas Amazônico, Cerrado, Mata Atlântica e a condição geral do rebanho (AGUIAR, 2013). A Figura 11 ilustra o mapa das pastagens amostradas.

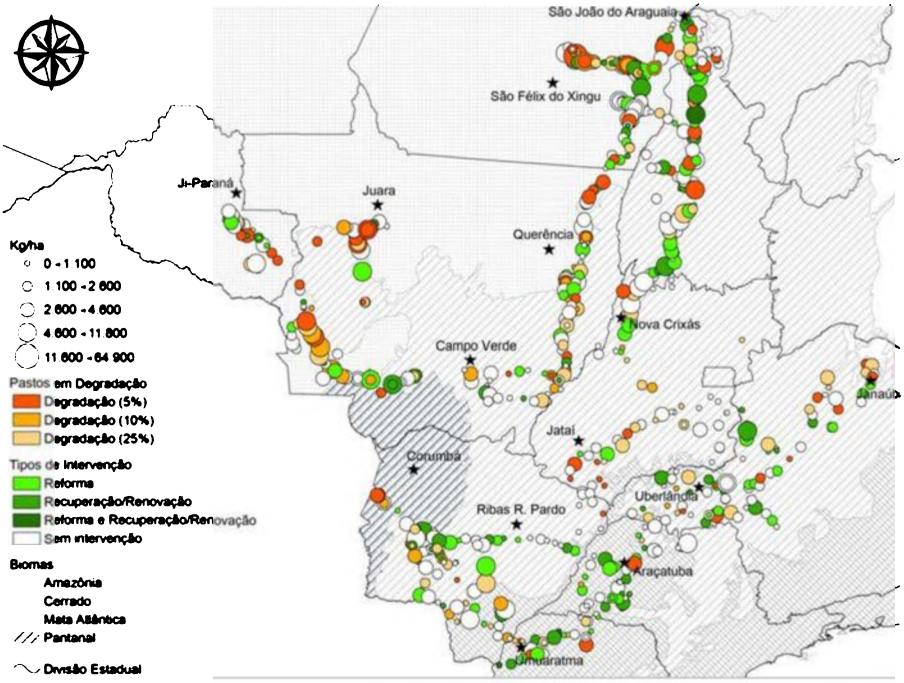
Estudos proposto por Aguiar (2013) apontam que a região que compreende o bioma Pantanal apresentou o maior percentual de pastagens em degradação embora neste bioma a amostragem fosse menos intensa. Já entre os biomas com maior número de pastagens amostradas, a Amazônia teve o maior percentual de pastagens em processo de degradação, com 34,5%, seguido pelo Cerrado, com 29,6%, e pela Mata Atlântica, com 18,9%.

1. Para detalhes veja www.geodegrade.cnpm.embrapa.br
2. Para detalhes veja www.rallydapecuaria.com.br

Digitized by Google

Na região Centro-Oeste, onde predomina o bioma Cerrado, a causa mais frequente de degradação das pastagens é o uso sistemático de animais que excedem a capacidade de recuperação da pastagem, ocasionando desfolhação e perda de nutrientes do solo, agravada pela ausência de manejo adequado para a recuperação das pastagens (FERREIRA *et al.*, 2012; MACEDO, 1997; OLIVEIRA *et al.*, 2004 *apud* AGUIAR, 2013).

Figura 11 - Pastagens amostradas em processo de degradação e com intervenções



Fonte: AGUIAR (2013)

Já na região Norte, onde predomina o bioma Amazônia, a maior causa de degradação da pastagem é a mudança de composição biológica devido ao processo de sucessão secundária, isto é, a mudança reduz a capacidade de a pastagem produzir economicamente devido a competição com as plantas invasoras (ASNER *et al.*, 2004; DIAS-FILHO *et al.*, 2001; TOWNSEND *et al.*, 2009 *apud* AGUIAR, 2013). Pastagens que tiveram problema no

Digitized by Google

estabelecimento ou que foram atacadas por insetos são mais suscetíveis a esse tipo de problema, chamado de degradação agrícola (DIAS FILHO, 2011).

Segundo Assad (2013) o Tocantins é um dos seis estados brasileiros que têm alguma aderência ao Plano ABC, isto é, apresenta nas suas políticas de mudança do clima, algum tipo de coesão com o plano. A Tabela 15 apresenta as metas estaduais em relação à abrangência territorial em milhões de hectares, o potencial de mitigação e a participação do estado na redução das emissões nacionais, para cada uma das ações propostas pelo plano ABC. Conforme a Tabela 16 observa-se que entre os compromissos estabelecidos pelo estado do Tocantins a área de florestas plantadas é o que possui maior destaque.

Tabela 16 - Processos Tecnológicos, metas estaduais até 2020, potenciais de mitigação e percentual em relação à meta nacional

Processo Tecnológico	Compromisso (milhões de ha)	Potencial de Mitigação (milhões Mg CO_2 eq)	Participação (%)
Recuperação de Pastagens Degradadas[1]	1,2	6,4 a 8	8%
Integração Lavoura – Pecuária – Florestas[2]	0,2	1,1 a 1,3	6%
Sistema de Plantio Direto (SPD)	0,4	0,7 a 0,9	4%
Fixação Biológica de Nitrogênio	0,2	0,3	3%
Florestas Plantadas	0,3	-	11%
Tratamento de dejetos animais	0,03	0,04	1%
Total	-	**8,5 a 10,5**	

[1] Por meio de manejo adequado e adubação
[2] Incluindo os sistemas agroflorestais (SAFs)
Fonte: BRASIL (2014)

Com relação à distribuição regional dos recursos do programa ABC e a participação do estado do Tocantins, dos 1.273 contratos assinados na região Norte para o programa ABC no ano safra 2013/2014, 682 (53%) foram assinados pelo Tocantins.

Digitized by Google

CAPÍTULO 4:

PROCEDIMENTOS METODOLÓGICOS

4.1 Métodos

Neste item serão abordados os métodos utilizados para a realização desta pesquisa, os quais foram:

- *Exploratório:* Por ser necessário ter uma visão panorâmica ou mais abrangente do setor de florestas plantadas no Brasil e no Estado do Tocantins, ao mesmo tempo, saber dados técnicos e especificidades sobre as regiões de estudo selecionadas.
- *Descritivo:* Por estabelecer relações entre as diversas variáveis técnicas e ambientais. Este tipo de pesquisa visa identificar estruturas, formas, funções e contextos.
- *Explicativo:* Busca mostrar as contribuições da classificação de áreas com reflorestamentos por meio de métodos não destrutivos realizados de forma rápida e com baixo custo.
- *Estatísticos:* Este estudo utilizou-se da técnica de mineração de dados, mais especificamente para os classificadores de árvore de decisão.

Para dar continuidade a este trabalho utilizaram-se também os seguintes tipos de pesquisa: a) Bibliográfica; d) Documental e Estudo de caso. Com base no relato acima, se observa que este trabalho segue uma natureza aplicada, pois propõe um estudo que pretende classificar áreas com florestas plantadas em meio a outros tipos de cobertura do solo, no estado do Tocantins, em imagens de sensoriamento remoto.

Os procedimentos de coleta dos dados utilizados, tanto para os dados qualitativos, quanto para os dados quantitativos foram do tipo secundário em sua maioria, mas não se isentando de ter usado

Digitized by Google

alguns do tipo primário.

O estudo de caso é o estado do Tocantins e para a obtenção de dados gerais e específicos de seus municípios foi pesquisado em sites do IBGE, ADAPEC, SEAGRO, SEMADE e outros.

Também foram realizadas entrevistas não estruturadas com profissionais sobre a matéria do trabalho, em órgãos estaduais como, ADAPEC, IBGE e SEAGRO. O objetivo das entrevistas foi obter informações sobre o setor de florestas plantadas, com possível finalidade energética.

4.1.1 Dados de sensoriamento remoto utilizados no trabalho

Foram utilizados na elaboração deste trabalho dados da plataforma Landsat 5 TM, (bandas TM1, TM2, TM3, TM4, TM5, TM7) obtidos junto ao Catálogo de Imagens disponibilizado no sítio do INPE na internet. Foram utilizadas as cenas 222/64, 222/65 e 223/64 conforme Tabela 17 onde são apresentadas as informações relativas aos horários de passagem do satélite, ângulo de incidência do sol e finalidade da imagem.

Tabela 17 - Plataformas e sensores utilizados no trabalho

Órbita Ponto	Data e hora do imageamento	Ângulo em graus	Justificativa do uso da imagem
222/64	17/09/2011 às 13h06	64,24º	Imagens utilizadas para identificação de florestas plantadas em meio a outros tipos de cobertura de solo
222/65	14/09/2010 às 13h29	59,45º	
223/64	07/08/2011 às 13h12	51,89º	

Fonte: Elaboração própria para a dissertação

Na etapa processamento dos dados do sensoriamento remoto utilizou-se o sistema de informações geográficas ENVI (*Environment for Visualising Images*), na versão 4.7. A respectiva ferramenta é amplamente utilizada no processamento de imagens de satélite e permitiu a visualização, o registro, as conversões de datum e o recorte das imagens Landsat/TM.

Para as análises estatísticas foi utilizado o software R (*Funda-*

Digitized by Google

tion para Statistical Computing), na versão, 3.0.1. Por meio desta ferramenta realizou-se a construção do plano amostral, o processo de seleção aleatória de municípios e os resultados utilizados nas análises dos dados desse trabalho.

Para a construção dos modelos de árvore de decisão, extração de regras e avaliação dos resultados da classificação utilizou-se a ferramenta WEKA – (*Waikato Environment for Knowledge Analisys*).

O WEKA foi escolhido devido a sua praticidade de utilização e pelo fato de ser um software livre, estando disponível em: http://www.cs.waikato.ac.nz/ml/weka.

4.1.2 Áreas de estudo

A área de estudo está localizada no estado do Tocantins, nas regiões geográficas conhecidas como Bico do Papagaio e Araguaína. Estas regiões foram selecionadas devido a grande concentração de florestas plantadas e ao fato de boa parte da destinação da madeira visar à obtenção de energia. Os municípios selecionados são apresentados na Tabela 18.

Tabela 18 - Áreas de estudo

Área de estudo	Coordenadas Geográficas	Datum	Cidades incluídas
Tocantinópolis	Lat 04º 52' 51,10'' a 06º 41' 26,21'' Lon 47º 57' 59.76'' a 46º 35' 21,12''	SAD-69	São Miguel do Tocantins São Bento do Tocantins
Araguatins	Lat 04º 52' 49,40'' a 06º 41' 24,07'' Lon 49º 30' 09.36'' a 48º 08' 00,96''	SAD-69	Araguatins
Araguaia	Lat 06º 19' 34,39'' a 08º 08' 10,39'' Lon 48º 16' 42.24'' a 46º 53' 47,40''	SAD-69	Babaçulândia Darcinópolis Wanderlândia

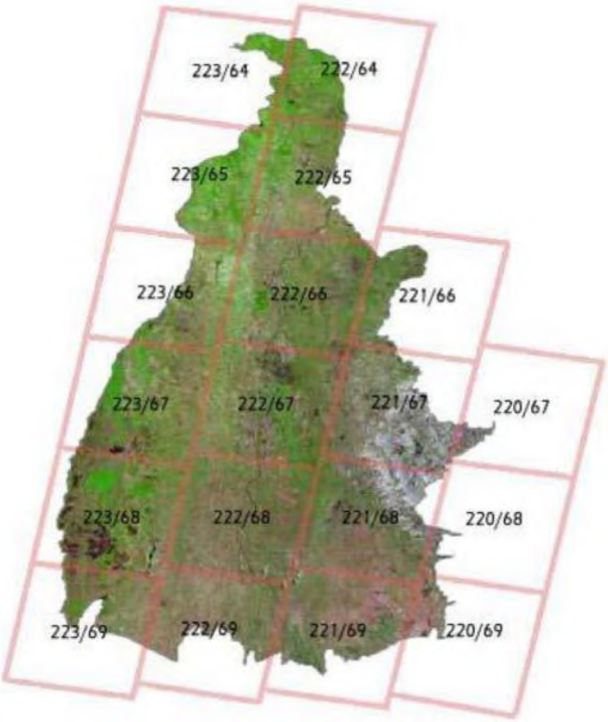
Fonte: Elaboração própria da dissertação

4.2 Seleção das áreas de estudo (Fase 1)

Este estudo abrangeu áreas com florestas plantadas no estado do Tocantins. A grade com a junção das cenas que podem ser inclusas no plano amostral deste estudo é apresentada na Figura 12.

Digitized by Google

Figura 12 - Grade das cenas Landsat 5/TM, Estado do Tocantins



Fonte: Elaboração própria para a dissertação

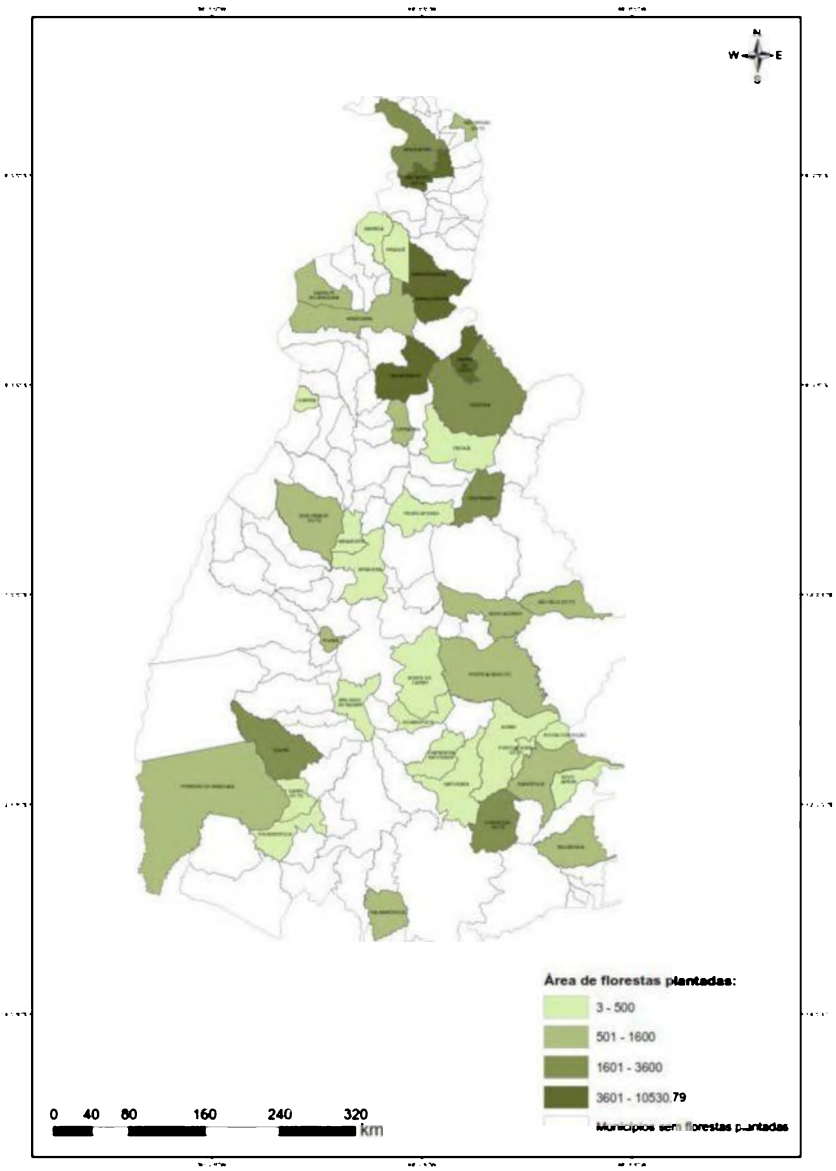
Na grade são ilustradas as cenas que cobrem o estado do Tocantins sob a malha Landsat 5TM. A malha disponibiliza informações sobre a posição das órbitas e pontos, permitindo que as imagens sejam organizadas e posteriormente localizadas.

4.2.1 Estimativa das áreas com florestas plantadas e o plano amostral

A partir dos dados obtidos junto a ADAPEC e para garantir a representatividade dos resultados optou-se por fazer uma amostragem estatística dos municípios com florestas plantadas, por meio da técnica de amostragem estratificada onde a os estratos são cenas do satélite Landsat 5. Inicialmente foi elaborado um mapa temático com a área plantada por município no Estado do Tocantins, para o ano de 2011/12, apresentado na Figura 13.

Digitized by Google

Figura 13 – Área plantada (em ha) com florestas por município no estado em 2011/12.



Fonte: Elaboração própria para a dissertação

Esta etapa foi realizada no software ArcGIS, que possui ferramentas apropriadas para a elaboração de layouts dos mapas. De acordo como a Figura 13 a concentração das áreas com florestas

Digitized by Google

plantadas cadastradas ocorre principalmente na região Norte e com menor intensidade na parte Sul do Estado.

No Norte do Estado os municípios de maior destaque, em área plantada são: São Bento do Tocantins, Wanderlândia, Babaçulândia, Palmeirante e Barra do Ouro. Já na região sul destacam-se os municípios de Dueré e Conceição do Tocantins, com maiores plantios florestais. A ADAPEC - Agência de Defesa Agropecuária do Estado do Tocantins forneceu a este pesquisador uma lista com os 30 municípios com maior destaque no cultivo de florestas plantadas do Estado do Tocantins, conforme o ANEXO I deste trabalho.

Tabela 19 - Municípios com maior área de florestas plantadas no estado do Tocantins em 2011/2012

Município	Área Total cultivada em (há)	Cena Landsat *TM* 5 (estrato)
Lagoa de São Bento	10530	222/64
Barra do Ouro	6883	222/65
Wanderlância	5962	222/65
Palmeirante	5234	222/65
Babaçulandia	4159	222/65
Dueré	3500	223/68
Araguatins	3498	223/64
Goiatins	3336	222/65
Centenário	1803	222/66
Conceição do Tocantins	1784	221/68
Araguaina	1148	222/65
Dianópolis	1012	221/68
Novo Acordo	906	222/67
Pugmil	900	222/67
Ponte Alta do Tocantins	742	221/67
Tupiratins	665	222/66
Taguatinga	656	221/69
Formoso do Araguaia	639	223/68
São Félix do Tocantins	568	221/67
Dois irmãos	534	223/66
Figueirópolis	460	222/68
Miracema	370	222/67
Miranorte	265	222/67
Xambioá	200	223/64
Novo Jardim	120	221/68
Itacajá	100	222/66
Couto Magalhães	94	223/66

Fonte: Dados da ADAPEC (2014). Levantamento relativo ao ano de 2011/2012
Elaboração própria

Digitized by Google

A base fornecida continha também os imóveis responsáveis pelos plantios, suas respectivas coordenadas geográficas (latitude e longitude) e a área utilizada em cada propriedade em hectares. Os 30 municípios listados na Tabela 19 são responsáveis por quase 90% da área com florestas plantadas no Estado. Estes municípios estão distribuídos em onze das 19 cenas do Satélite Landsat 5/TM que cobrem o estado do Tocantins, conforme ilustra a Figura 14. Nas cenas destacadas em cor laranja estão contidos os municípios listados na Tabela 19.

Figura 14 - Cenas relativas aos 30 municípios com maiores plantios florestais no Tocantins

Fonte: Elaboração própria

Após a composição das cenas disponíveis na Figura 14, foi executado o processo de obtenção dos extratos para as áreas de estudo, buscando obter a melhor representatividade possível dos dados. Optou-se por utilizar como estrato, as cenas do satélite Landsat 5/TM, conforme ilustra a Tabela 20.

Digitized by Google

Tabela 20 - Cidades com maior área de florestas plantadas distribuídas por cena Landsat 5 TM

Cena Landsat 5 TM	Município
221/67	Ponte Alta do Tocantins
	São Félix do Tocantins
221/68	Conceição do Tocantins
	Dianópolis
	Novo Jardim
221/69	Taguatinga
222/64	Lagoa de São Bento
222/65	Araguaína
	Babaçulandia
	Barra do Ouro
	Goiatins
	Palmeirante
	Wanderlância
222/66	Centenário
	Itacajá
	Tupiratins
222/67	Miracema
	Miranorte
	Novo Acordo
	Pugmil
222/68	Figueirópolis
223/64	Araguatins
	Xambioá
223/66	Couto Magalhães
	Dois irmãos
223/68	Dueré
	Formoso do Araguaia

Fonte: Elaboração própria

A partir da estratificação por cena, foram selecionados os municípios para compor a amostra inicial. A variável utilizada no cálculo da média e da variância foi a área total com plantios florestais no município. Utilizou-se a técnica de amostragem aleatória estratificada e os parâmetros utilizados no cálculo do tamanho da amostra são apresentados na Tabela 21.

Digitized by Google

Tabela 21 - Parâmetros utilizados no cálculo do tamanho da amostra dos municípios de estudo

Estrato	Média da área com floresta plantada (ha)	Número de muninípios no estrato
221/67	655,0	2
221/68	972,0	3
221/69	656,0	1
222/64	10530,0	1
222/65	4453,7	6
222/66	856,0	3
222/67	610,3	4
222/68	460,0	1
223/64	1849,0	2
223/66	314,0	2
223/68	2069,5	2

Fonte: elaboração própria

O processo de amostragem resultou na seleção de 15 municípios para compor a amostra inicial. Por meio de um sorteio aleatório simples dos municípios dentro de cada um dos estrados (cenas do satélite Landsat 5) selecionou-se pelo menos um município de cada um dos 11 estratos. Os municípios inicialmente selecionados estão listados na Tabela 22.

Tabela 22 – Relação de municípios selecionados por amostragem

Cena Landsat 5/TM	Município
221/67	São Félix do Tocantins
221/68	Conceição do Tocantins
221/69	Taguatinga
222/64	Lagoa de São Bento
222/65	Babaçulandia
	Darcinópolis
	Wanderlância
222/66	Centenário
222/67	Miracema
	Pugmil
222/68	Figueirópolis
223/64	Araguatins
	Xambioá
223/66	Dois irmãos
223/68	Dueré

Fonte: Elaboração própria

Digitized by Google

A seguir com os respectivos dados, iniciou-se a fase da busca por imagens nos repositórios utilizados. O principal repositório de imagens Landsat 5/TM utilizado no trabalho foi o Banco de Imagens da DGI/INPE. Na interface de busca do sistema foram inseridos os parâmetros de interesse, como por exemplo, os nomes dos municípios, cobertura máxima de nuvens, entre outros. Assim as imagens de interesse foram baixadas para por meio de um sistema de FTP (File Transfer Protocol).

O repositório de imagens foi consultado diversas vezes a fim de encontrar imagens sem falhas nos registros de NDs dos pixels, imagens com cobertura mínima de nuvens e imagens correspondentes às áreas selecionadas para o estudo.

Em seguida iniciou-se o processo de identificação das áreas com florestas plantadas nas imagens adquiridas.

Apesar da evolução crescente das áreas com florestas plantadas no estado, as áreas ocupadas com estes plantios, nos municípios inseridos no estudo são pequenas em relação à área total disponível para agricultura, o que não permite que os plantios sejam visualizados facilmente na imagem. Para auxiliar as buscas por áreas com reflorestamentos dentro das imagens Landsat 5, foram utilizadas algumas ferramentas e dados, como a base de dados com a localização dos plantios fornecida pela ADAPEC e a ferramenta para visualização de séries temporais derivadas de imagens de sensoriamento remoto (series view), desenvolvido pelo laboratório de Sensoriamento Remoto Aplicado à Agricultura e Floresta. Estas ferramentas foram utilizadas para a identificação de outros tipos de alvos na superfície terrestre ou possíveis mudanças de cobertura com influencia no estudo.

Os outros alvos inseridos no estudo foram área urbana, floresta ou vegetação nativa, corpos d'água como rios e lagos e áreas contendo solo exposto.

O *series view* permite a visualização instantânea e análise de

Digitized by Google

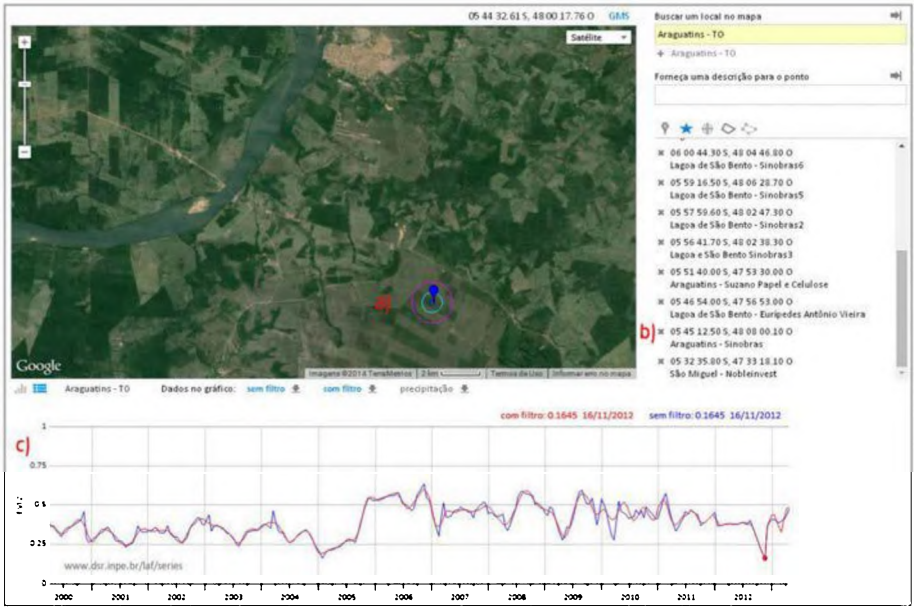
séries temporais de produtos MODIS integradas a um globo virtual, como o GoogleMaps, e pode ser utilizada como suporte a problemas de uso e cobertura da terra (FREITAS *et al.*, 2011). A ferramenta permite o acesso gratuito a uma série temporal de 10 anos de dados MODIS em qualquer coordenada geográfica da América do Sul e pode ser acessada no endereço https://www.dsr.inpe.br/laf/series/index.php.

Para ilustrar o processo de busca e visualização de áreas com florestas plantadas, a Figura 15 apresenta uma área plantada com eucalipto próxima ao município de Araguatins-TO. Na Figura 15 é possível observar a delimitação dos plantios em torno do ponto selecionado. Na Figura 15-b observa-se a relação de municípios selecionados para validação das áreas plantadas e das coordenadas dos imóveis com reflorestamentos. A Figura 15-c é apresenta o comportamento do índice de vegetação EVI2 ao longo do tempo. O comportamento da série é semelhante ao do ciclo de curta rotação de plantios de eucalipto, onde o corte ocorre por volta dos sete anos de idade. O índice detecta um ganho de biomassa em 2004 e os valores de EVI2 aumentam. No período de 2005 a 2011 a resposta espectral é estável durante todo o período. No ano de 2012 a queda brusca no valor do índice sugere a ocorrência de um corte por volta dos 7 anos. A linha azul na série representa os dados originais e a série vermelha é a filtrada por meio da técnica conhecida como transformada *wavelets*, que visa a eliminação de possíveis ruídos.

Utilizando estes materiais auxiliares e inspeção visual foi possível identificar as regiões de interesse. Mesmo possuindo data defasada, as imagens possibilitaram a identificação de diferentes coberturas do solo, como áreas urbanas, lagos e rios, estradas, vegetação nativa e florestas plantadas.

Digitized by Google

Figura 15: a) Ponto referente a área plantada no município de Araguatins, b) Listagem de imóveis com plantios; c) Série temporal do indice de vegetação EVI-2 para a coordenada selecionada



Fonte: Elaboração própria. Adaptado do site www.dsr.inpe.br/laf/series

4.2.2 Lista final de municípios

Após realizar buscas por imagens Landsat 5/TM, verificou-se que muitas imagens apresentavam incompatibilidades radiométrica e cobertura de nuvem. Por isso, optou-se por reduzir o número de municípios em estudo e o número de cenas Landsat 5/TM cobrindo o estado. Para garantir a representatividade dos dados, levando em consideração o porte dos reflorestamentos, o estudo foi delimitado as cidades cobertas pelas cenas Landsat 5/TM de órbita/ponto 222/64, 222/65 e 223/64. Estas cenas estão entre as poucas imagens livres de nuvens no período de 2010/2011. A Tabela 23 contém os nomes dos municípios selecionados, após a aplicação da amostragem aleatória simples na lista de municípios contidos na cena 222/64, 222/65 e 223/64. A aplicação do procedimento de amostragem aleatória simples dentro da cena ou estrato, foi possível porque os municípios contidos nelas possuem áreas

Digitized by Google

expressivas de reflorestamentos.

Tabela 23 - Lista final de municípios e respectivas áreas com plantios florestais inseridas no trabalho

Cidade	Área do reflorestamento em (ha)
Araguatins	13.145,0
São Bento do Tocantins	7.100,0
Wanderlândia	5.051,0
Babaçulândia	4.159,0
Darcinópolis	2.511,0
São Miguel do Tocantins	1.480

Fonte: STCP, 2013

4.3 Processamento das imagens (Fase 2)

Após a definição das áreas de interesse, as imagens correspondentes às três cenas em estudo foram processadas a fim de realizar a correção radiométrica e corrigir imperfeições no registro dos pixels da imagem e a posterior conversão dos números digitais (DN) registrados nos pixels das imagens em valores de refletância.

4.3.1 Correção radiométrica

Nesta etapa foi realizada a verificação da qualidade de cada uma das imagens utilizadas por meio de inspeção visual a fim de encontrar falhas ou diferenças radiométricas entre as imagens Landsat 5/TM. Após tal inspeção visual das imagens foram analisadas as estatísticas das imagens, representadas pelos histogramas de níveis de cinza, a fim de encontrar possíveis erros nos registros dos pixels das imagens. Como não foram encontradas falhas nas imagens correspondentes às três cenas em estudo, nenhum método ou correção foi utilizado.

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

Para a realização da conversão para valores de reflectância

Digitized by Google

utilizou-se o método proposto por Chander e Markhan (2003), usando os coeficientes de calibração listados na Tabela 24. Este procedimento de calibração foi aplicado em cada uma das imagens utilizadas. Os coeficientes foram aplicados aos dados das imagens Landsat 5, utilizando-se o módulo de calibração do sistema de processamento de imagens ENVI 4.7. A Equação utilizada na conversão é apresentada a seguir:

$$L = Lmim + \frac{ND*(Lmax - Lmin)}{255} \qquad 3.6$$

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;
- *Lmax* é o valor máximo do coeficiente de pós-calibração da banda;
- *Lmin* é o valor mínimo do coeficiente de pós-calibração da banda.

Tabela 24 - Pós-calibração sensor Landsat 5/TM

Banda	Lmin	Lmax	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

Digitized by Google

4.4 Preparação dos dados (Fase 3)

Nesta etapa realizou-se a preparação dos dados de sensoriamento remoto por meio da extração de informações das imagens. Foram aplicadas transformações nos dados, por meio de cálculo de índices de vegetação.

Os índices de vegetação são caracterizados por combinações algébricas entre as bandas espectrais das imagens. Em geral, às bandas do vermelho e infra-vermelho próximo são as mais utilizadas no cálculo dos índices de vegetação.

4.4.1 Extração dos dados relativos às regiões de interesse

Após a identificação das áreas de interesse, as mesmas foram marcadas em diversas cores, por meio da utilização da ferramenta "*ROI Tool*" do sistema de processamento de imagens ENVI 4.7. Posteriormente realizou-se a extração dos dados relativos às regiões de interesse. Os recortes das imagens foram então exportados para arquivos no formato TXT. O processamento foi repetido para cada uma das imagens, inclusive as imagens de textura. O resultado final desse procedimento consistiu diversos arquivos TXT contendo informações sobre a reflectância dos alvos selecionados.

4.4.2 Transformação dos dados da imagem para o formato tabular

Depois que os arquivos em formato TXT relativos as regiões de interesse foram todos gerados, iniciou-se a organização dos dados nos arquivos. Os arquivos foram convertidos em planilhas excel e as classes em estudo foram atribuidos a cada linha, isto é, cada linha da planilha passou a representar um pixel.

4.4.3 Consolidação dos dados em arquivos individuais

Os multiplos arquivos gerados por meio da ferramenta ROI

Digitized by Google

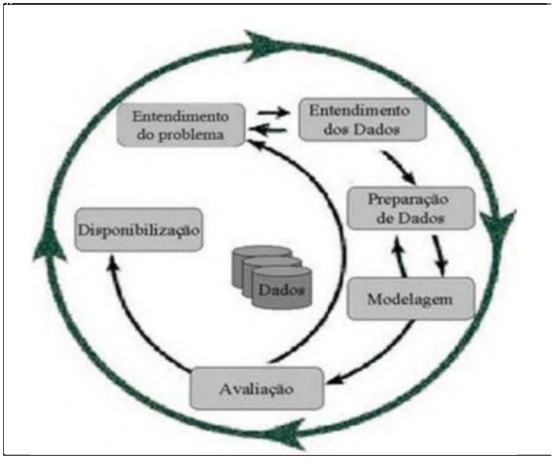
tool do sistema ENVI, foram reunidos em uma unica planilha e então realizou-se a conversão das planilhas para arquivos do tipo ARFF (Atribute Relation File Information), que são arquivos específicos para armazenamento de dados no sistema de dados WEKA. O arquivo denominado ARFF utilizado continha 52.847 observações.

Com estes arquivos contendo os 60 atributos em estudo (60 atributos preditivos e um atributo resposta), iniciou-se a aplicação de técnicas de mineração de dados , tanto para a seleção de atributos, como para a classificação de alvos em estudo.

4.5 Mineração de dados (Fase 4)

O trabalho de mineração de dados foi realizado tendo em vista o seguinte cenário: Construção de um modelo de classificação de áreas com plantios florestais em meio a outros tipos de cobertura de solo. A partir dos arquivos consolidados, foi iniciado o trabalho de mineração de dados, em particular, a tarefa de classificação foi. Nesta etapa foi empregada a metodologia CRISP--DM (*Cross Industry Standard Process for Data Mining*) criada por CHAPMAN et al., (2000) e apresentada na Figura 16.

Figura 16: CRISP-DM



Fonte: MEIRA (2008)

Digitized by Google

O modelo do processo de mineração de dados CRISP-DM, apresentado na Figura 16, consiste na divisão do um projeto de mineração de dados em seis fases: entendimento do problema, entendimento dos dados, preparação dos dados, modeloagem, avaliação e distribuição.

4.5.1 Definição dos atributos e entendimento dos dados

O ponto inicial desta fase do trabalho foi à definição e descrição de cada um dos atributos utilizados na construção do modelo de classificação.

A Tabela 25 lista a definição e descrição do atributo alvo, isto é, variável resposta empregada nos modelos de classificação do cenário em estudo.

Tabela 25 - Atributo resposta para o Cenário em estudo composto por cinco classes

Variável alvo Classe de interesse	Valores que a variável pode assumir	Descrição
CLASSE_ALVO_INTERESSE	Área_urbana Solo_exposto Reflorestamento Corpos_dagua Vegetaca_nativa	Variável qualitativa que poderia assumir até cinco valores nominais diferentes. Cada um dos valores correspondentes a um tipo de cobertura de solo na imagem estudada

Fonte: Elaboração própria

A Tabela 26 apresenta os principais atributos preditores utilizados no modelo e a descrição detalhada de cada uma deles. Estes atributos são relativos à resposta espectral dos alvos em estudo nas diferentes bandas do espectro eletromagnético visível.

Tabela 26 - Atributos relativos à resposta espectral dos alvos em estudo nas diferentes bandas do espectro eletromagnético

Variável	Descrição
B1	Fator de refletância Bidirecional (FRB) de superfície para a banda TM1 (azul)
B2	FRB de superfície para a banda TM2 (verde)
B3	FRB de superfície para a banda TM3 (vermelho)
B4	FRB de superfície para a banda TM4 (infravermelho próximo)
B5	FRB de superfície para a banda TM5 (infravermelho médio)
B7	FRB de superfície para a banda TM2 (infravermelho distante)

Fonte: Elaboração própria

Digitized by Google

A Tabela 27 apresenta os índices de vegetação utilizados nesse trabalho. Mais detalhes sobre os índices de vegetação podem ser obtidos no Tópico 2.1.3. Estes índices foram considerados no cenário em estudo, visto que são atributos relevantes para o estudo da interação da radiação eletromagnética com os dosséis vegetais.

Tabela 27 – Índices de vegetação utilizados pelo modelo de classificação

Variável	Descrição
NDVI	Índice de vegetação por diferença normalizada
PVI	Índice de vegetação perpendicular
EVI	Índice de vegetação melhorado
RVI	Índice de vegetação pela razão de bandas
SAVI	Índice de vegetação ajustado pelo solo

Fonte: Elaboração própria

Na tabela 28, observam-se os atributos de textura considerados no trabalho. Detalhes sobre a definição de atributos de textura podem ser observados no Tópico 2.1.5 referente à extração de informação de imagens de satélite. Os atributos de textura foram extraídos para todas as bandas. O processo de extração de medidas de textura gerou 48 novos atributos, mais especificamente, 48 novas imagens. Isto é, para cada uma das 6 bandas, foram gerados 8 atributos de textura.

Tabela 28 – Definição dos atributos de textura utilizados na classificação

Variável	Descrição
Média	Média dos valores observados dos pixels vizinhos
Variância	Variabilidade dos valores observados dos pixels vizinhos
Homogeneidade	Uma medida de uniformidade local dos valores dos pixels vizinhos
Contraste	Uma medida de variação local dos valores dos pixels vizinhos
Dissimilaridade	Uma medida de diferenciação entre os pixels vizinhos
Entropia	É uma medida de aleatoriedade dos pixels vizinhos que assume baixos valores para imagens suaves e homogêneas
Segundo Momento	Uma medida de suavidade local
Correlação	Uma medida de associação entre os pixels vizinhos

Fonte: Elaboração própria

Digitized by Google

A janela de processamento considerada teve tamanho nove, isto é, a vizinhança processada possuía distancia de em pixel na horizontal e um pixel na vertical. Estes atributos foram gerados para cada uma das seis bandas diferentes do sensor Landsat 5/TM.

4.5.2 Seleção de atributos

O conjunto de dados resultante do processo de preparação de dados possuía 60 atributos (59 atributos preditores e um atributo resposta definindo as classes de interesse). A definição dos 60 atributos pode ser vista no Tópico 4.5.1. Com o objetivo de entender a contribuição específica de cada atributo, maximizar as dissimilaridades e minimizar a similaridade entre atributos, foram empregados métodos de seleção listados na Tabela 29.

Tabela 29 - Métodos de seleção de atributos em relação ao mérito individual

Método de seleção de atributo	Descrição
Ganho de Informação	H(classe) - H(classe |atributo)
Razão de Ganho de Informação	H(classe) - H(classe |atributo) / H(atributo)
Qui-quadrado	Usa uma medida estatística de distância da variável em relação às classes
weka.attributeSelection.CfsSubsetEval	Avalia a capacidade preditiva individual de cada atributo e a redundância entre atributos

Fonte: Elaboração própria

4.5.3 Aplicação do modelo de classificação

Esta etapa consistiu na classificação dos dados, por meio do método de Árvore de Decisão Binária. O algoritmo de indução utilizado 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. Os parâmetros utilizados na construção dos modelos de

Digitized by Google

classificação são apresentados na Tabela 30.

Tabela 30: Parâmetros do modelo de árvore de decisão

Parâmetro	Valor
Divisão	Binária
Fator de confiança	0,25
Número mínimo de objetos por folha/regra	5,10,20,30,40,50,60,70,...100,200,...500,600, 700
Poda	Sim (Substituição de sub-árvore)
Precisão de Laplace	Não

Fonte: Elaboração própria

Os diferentes números mínimos de objetos por folha mencionados na Tabela 29 estão associados aos diferentes tipos de poda que podem ser utilizados na construção das árvores de decisão. Quanto maior o número mínimo de pixels por folha, na árvore de decisão, mais abrangente e menos específica é a cobertura da regra de classificação.

Foram utilizadas 6 configurações diferentes de atributos para construção dos modelos que foram posteriormente comparados. Cada uma das configurações foi resultado da aplicação dos métodos de seleção de atributos citados no tópico 4.5.2 e de outros três conjuntos listados na Tabela 31.

Tabela 31: Seleção de subconjuntos mais adequados para a classificação

Subconjunto de atributos frequentemente utilizados pelos especialistas em sensoriamento remoto	Descrição do subconjunto de atributos
Somente as Bandas (B1, B2, B3, B4, B5 e B7)	Nessa configuração estão somente as bandas do sensor TM
Bandas e índices de vegetação (B1, B2, B3, B4, B5, B7, NDVI, RVI, PVI, SAVI)	Todas as bandas e mais cinco índices de vegetação
Todos os atributos	O modelo foi construído com todos os atributos no estudo

Fonte: Elaboração própria

Digitized by Google

4.5.4 Avaliação dos resultados da classificação

Após a etapa de seleção dos diferentes conjuntos de atributos mais relevantes para o estudo, foram construídos os modelos de classificação baseados no método de aprendizagem por árvore de decisão para o cenário em estudo. Ao todo foram utilizados 6 configurações de atributos das quais 3 foram resultantes de métodos de seleção de atributos e outras 3 que são tipicamente utilizadas por especialistas em sensoriamento remoto.

O método de avaliação da acurácia da classificação foi a análise da matriz de erros dos modelos e da estatística KAPPA (Tópico 2.5.3), que leva em consideração o número de acertos e o número de erros do modelo de classificação e fornece o desempenho geral da mesma. Também foram listadas as taxas de acertos absolutos e erro absoluto dos modelos, que embora não sejam o principal critério de avaliação, fornecem um bom indicativo de qualidade do modelo produzido.

Em seguida a análise da taxa de acerto, iniciou-se a análise das métricas da matriz de confusão ou matriz de erros (Tópico 2.5.2), para as classes de atributo do estudo. O objetivo principal desta análise foi identificar modelos que apresentassem uma boa acurácia e que fornecessem medidas de precisão e sensitividade satisfatórias do ponto de vista de um especialista do domínio.

Digitized by Google

CAPÍTULO 5:

RESULTADOS E DISCUSSÕES

Como mencionado no trabalho buscou-se avaliar o potencial das técnicas de mineração de dados na identificação de áreas cultivadas com florestas plantadas em meio a diferentes tipos de cobertura de solo: áreas urbanas, solo exposto, vegetação nativa e corpos d'água. Para isso foi aplicado o modelo de Árvore de Decisão Binária.

Para auxiliar na compreensão dos resultados, esta parte do trabalho foi dividida em quatro subitens: seleção de atributos, avaliação do modelo de melhor desempenho, avaliação de diferentes níveis de pré-poda para o modelo com melhor desempenho e considerações sobre os resultados obtidos.

5.1 Seleção inicial de atributos

A seleção de quais atributos seriam utilizados na análise se baseou em vários critérios, incluindo sua relevância para as metas do estudo e as restrições técnicas quanto à qualidade e volume dos dados. Conforme foi visto na revisão bibliográfica essa escolha tem como objetivo reduzir a dimensionalidade dos dados, melhorar o desempenho dos algoritmos, produzir modelos acurádos e permitir aos usuários foco nas variáveis mais importantes.

Nesse sentido, foram aplicados métodos de seleção de atributos aos dados do estudo e os resultados são apresentados nas Tabelas 32 e 33. A Tabela 32 refere-se a quatro métodos de seleção de atributos, considerando a contribuição individual de cada atributo

Digitized by Google

e a Tabela 33 apresenta os resultados da seleção 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.

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

Tabela 32: Avaliação de atributos em relação a contribuição específica para a classificação

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

Fonte: Elaboração própria

Digitized by Google

Tabela 33: Avaliação de subconjuntos de atributos frequentemente utilizados pela comunidade de sensoriamento remoto

Conjuntos frequentemente utilizados	Taxa de acerto do modelo	Estatística KAPPA	Número de Regras
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: Elaboração própria

5.2 Avaliação do modelo de classificação de maior acurácia

O melhor resultado de classificação foi obtido pelo conjunto de 6 atributos composto pelas bandas do espectro: Azul (B1), Verde (B2), Vermelho (B3), Infravermelho-próximo (B4), Infravermelho-médio (B5) e Infravermelho-distante (B7). Ao todo foram utilizados 52.847 registros, dos quais 34.876 (%66) foram utilizados para treinamento e 17.971 para testar a classificação. A Tabela 34 apresenta os resultados classificação dos diferentes alvos por meio da matriz de confusão.

Tabela 34: Matriz de confusão para o melhor modelo de classificação

		Classificado como				
		Área urbana	Vegetação Natural	Florestas plantadas	Solo exposto	Corpos d'água
Classe	Área urbana	[illegible]	1	116	328	0
	Vegetação Natural	6	[illegible]	106	0	0
	Florestas plantadas	63	77	[illegible]	91	0
	Solo exposto	160	0	162	[illegible]	0
	Corpos d'água	1	0	0	0	[illegible]

Fonte: Elaboração própria

Digitized by Google

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. As medidas relativas à qualidade do modelo estão listadas na Tabela 35.

Tabela 35: Métricas de qualidade do modelo extraídas da matriz de confusão

Classe	Sensitividade	Confiabilidade Positiva	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: Elaboração própria

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%.

A árvore de decisão gerada possui 57 folhas, também chamadas de regras, que são utilizadas pelo algoritmo classificador. A árvore de decisão, as regras e os resultados do modelo detalhado estão disponíveis no APÊNDICE I.

O número de regras obtido torna a interpretação demasiadamente difícil, pois para que a precisão do modelo fosse alcançada, foi necessário o particionamento dos dados em pedaços muito pequenos.

Digitized by Google

5.3 Avaliação de diferentes níveis de pré-poda para o modelo de melhor acurácia.

A poda em modelos de árvore de decisão é um ajuste com o objetivo de evitar overfitting, conforme explicado na revisão bibliográfica (Tópico 2.5.2). É 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 36 são apresentadas a Acurácia ou Taxa de acerto, a estatística Kappa e o número de regras geradas para níveis diferentes de pré-poda.

Tabela 36: 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: Elaboração própria

Os gráficos referentes a cada uma das colunas da Tabela 36 são apresentados a seguir nas Figuras 17, 18 e 19. Por meio da análise da Tabela 36 e das Figuras, 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

Digitized by Google

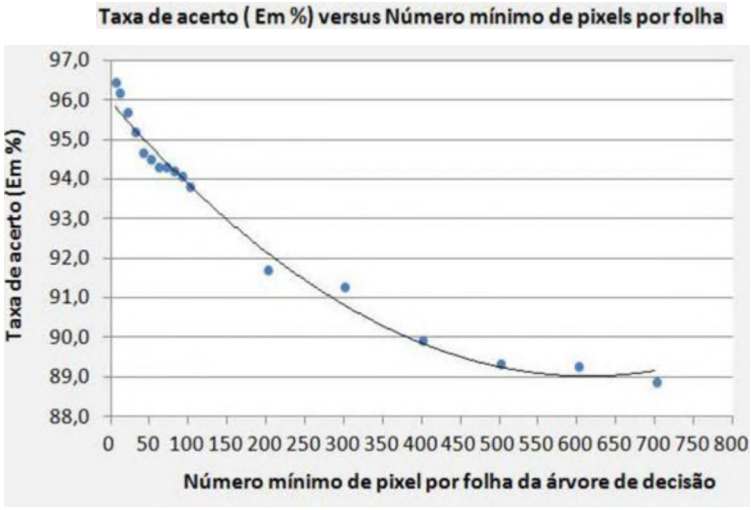
folha da árvore ou regra de decisão gerada.

Da análise da Figura 18, verifica-se que a estatística Kappa se mantém acima de 0,80, mesmo para um número mínimo de pixels, por folha, maior que quinhentos.

Ainda analisando-se a Tabela 36 e as Figuras 17 e 18, 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, conforme ilustrado na Figura 17.

Outra informação relevante da Tabela 36 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%.

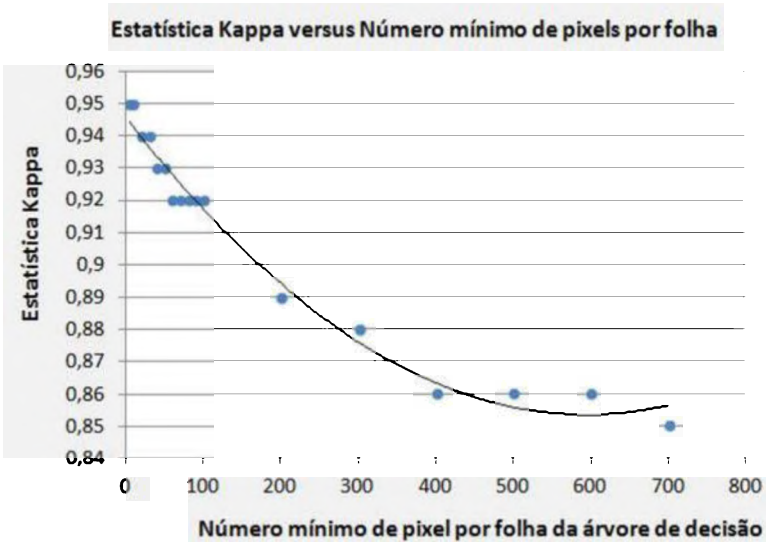
Figura 17: Taxa de acerto segundo o número mínimo de pixels por folha



Fonte: Elaboração própria

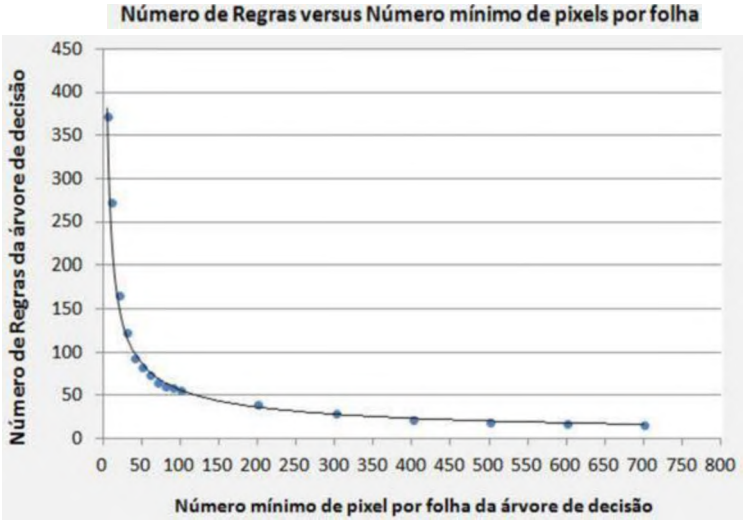
Digitized by Google

Figura 18: Estatística Kappa versus o número mínimo de pixels por folha



Fonte: Elaboração própria

Figura 19: Nº de regras da árvore de decisão versus o número mínimo de objetos por folha



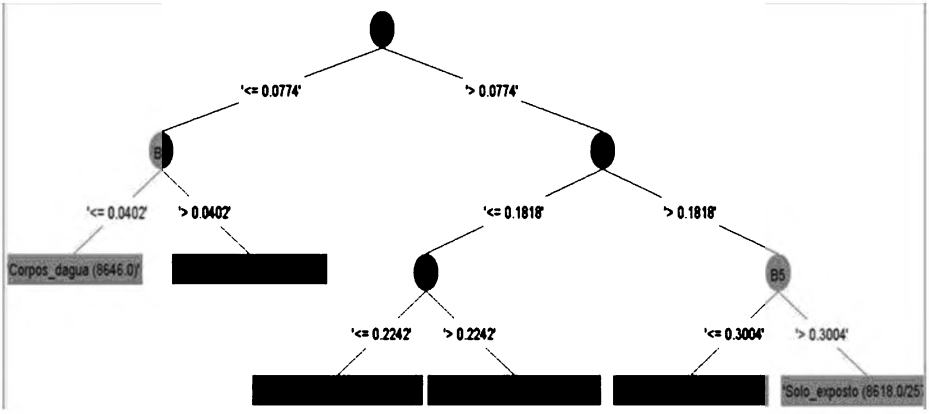
Fonte: Elaboração própria

Analisando-se a Figura 19 é possível verificar que o número de regras e consequentemente, o tamanho da árvore diminui con-

Digitized by Google

forme ocorre o aumento dos níveis de pré-poda. Dependendo da aplicação de interesse, poderia ser vantajoso para o analista de dados escolher modelos produzindo um número menor de regras em detrimento da acurácia do modelo. A Figura 19 pode auxiliar o analista a escolher um ponto de corte ótimo tendo em vista a otimização tanto da acurácia como do número de regras. Na Figura 20, somente a título de ilustração, observa-se a árvore de decisão gerada com nível de pré-poda igual a 2500 pixels por folha.

Figura 20: Árvore de decisão para um nível de pré-poda igual a 2500



Verifica-se que apenas 6 folhas possibilitaram uma taxa de acerto de 81%, conforme Figura 20. Também, verifica-se que apenas três atributos foram suficientes para gerar um modelo com acurácia superior a 80%. Os atributos são B4, B5 e B7.

A árvore apresentada na Figura 20, em princípio, não possui aplicação prática, mas permite a identificação atributos relevantes bem como descrições mais claras sobre a partição dos dados. A taxa de acerto dessa árvore é de pouco mais de 91%, de modo que em uma aplicação onde o tempo é importante, o analista de dados poderia utilizar este resultado sem com isto ter uma diminuição demasiada na acurácia geral da classificação.

Digitized by Google

5.4 Considerações sobre os resultados obtidos

Considerando o conjunto de atributos mais eficiente, segundo os métodos de seleção, o modelo obtido foi capaz de acertar mais de 93% dos casos. O resultado da principal medida de qualidade (estatística Kappa) obtida pelo 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 reflorestamentos em imagens.

Deve-se destacar também que, a diferença entre o resultado de classificação obtido pelo melhor conjunto de atributos e a classificação feita com todos os atributos do estudo é cerca de 0,22% e que os atributos de textura não foram selecionados por nenhum dos métodos de seleção de atributos o que sugere a baixa relevância de atributos de textura na classificação.

Digitized by Google

CONCLUSÕES E RECOMENDAÇÕES

A partir dos dados de localização, espacial e físico de florestas plantadas no TO, obtidos em pesquisa de campo junto a ADAPEC, foi possível a elaboração de um mapa temático apresentando a distribuição da área dos plantios florestais por município no Estado (Figura 13) e, dessa forma, identificar as áreas mais adequadas à análise no estado. Optou-se, então, por analisar imagens de satélite da região norte do estado do Tocantins, onde há maior disseminação de florestas plantadas.

As técnicas de mineração de dados utilizadas neste trabalho se mostraram 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, visto que os métodos de seleção de atributos identificaram subconjuntos reduzidos de atributos. 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.

O tamanho das árvores induzidas impossibilitou a análise das regras da classificação de maneira simples, devido ao número excessivo de nós e folhas gerados durante a aplicação algoritmo de classificação. Por essa razão, não foi possível extrair um conjunto

Digitized by Google

de regras compreensíveis. Uma maneira de resolver este problema seria obter árvores menores e mais compreensíveis por meio da utilização de diferentes níveis de pré-poda, mas sem perder muita precisão.

O subconjunto de atributos selecionados e as regras de classificação produzidas por esses atributos podem ser utilizados como base de conhecimento para um sistema especializado na identificação de florestas plantadas.

Os atributos de textura e índices de vegetação não trouxeram ganho no resultado da classificação realizada. Este resultado pode ser explicado pela baixa resolução espacial do sensor (30 metros) e pelas características semelhantes do dossel dos plantios das espécies em questão.

Então, a análise desenvolvida demonstra a necessidade de ampliar os estudos na área de sensoriamento remoto orbital. A caracterização do processo de descoberta do conhecimento em bancos de dados é fundamental para iniciativas atuais e futuras de monitoramento das florestas plantadas, ou seja, o planejamento desse tipo de produção e consequentemente, contribui no momento de tomada de decisão por parte de gestores do setor bioenergético.

Em função da experiência adquirida durante a realização da presente dissertação, como recomendações para estudos futuros, sugere-se:

- Aplicar este método em outras regiões com plantios florestais para validar o método sua adequação e limitações.
- Investigar os métodos utilizados neste trabalho em aplicações de monitoramento da biomassa e do estoque de carbono por área em uma determinada área com plantios florestais.
- Investigar a utilização de outras técnicas de mineração de dados, voltadas para a classificação, como a regressão logística e redes neurais.

Digitized by Google

REFERÊNCIAS BIBLIOGRÁFICAS

AGUIAR, D. A. **Séries temporais de imagens MODIS para avaliação de pastagens tropicais. 2013**. 167 p. (sid.inpe.br/mtc-m19/2013/08.30.14.33-TDI). Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2013. Disponível m: <http://urlib.net/8JMKD3MGP7W/3ENND62>. Acesso em: 24 jan. 2014.

ALVARES, C. A., MUNHOZ, J. S. B., STAPE, J. L., STAHL, J., ANDRADE, L., SANTOS, G., ... & LTDA–CAÇADOR, J. F. **Caracterização da dinâmica dos índices de vegetação NDVI e EVI em plantações de Pinus do Brasil**. *Anais XVI Simpósio Brasileiro de Sensoriamento Remoto-SBSR, Foz do Iguaçu, PR, Brasil, 13*.

AGRAWAL R.; ARNING A.; BOLLINGER T.; MEHTA M.; SHAFER J.; SRIKANT R. **The Quest Data Mining System, International Conference on Knowledge Discovery and Data Mining**, 2, 1996, Portland, Oregon, Aug. 1996.

ASSAD, E. D. **Agricultura de baixa emissão de carbono: a evolução de um novo paradigma**. São Paulo: [s.n.], 2013. Disponível em: <http://www.observatorioabc.com.br/ckeditor_assets/attachments/38/2013_06_28_relatorio_estudo_1_observatorio_abc.pdf>

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

BATISTA, G.T. AFFONSO, A. **Introdução ao geoprocessamento e ao sensoriamento remoto.** (Bacharelado em Agronomia). Universidade de Taubaté – UNITAU. 2002. Apostila.

BATISTA, M. H. ; HAERTEL, V. **Classificação Hierárquica Orientada a Objetos em Imagens de Alta Resolução Espacial Empregando Atributos Espaciais e Espectrais.** Dissertação de Mestrado. Universidade Federal do Paraná(UFPR). Porto Alegre-PR.2002.

BRASIL. Plano Setorial **de Mitigação e de Adaptação às Mudanças Climáticas para a Consolidação de uma Economia de Baixa Emissão de Carbono na Agricultura**, disponível em: <http://www.mma.gov.br/images/arquivo/80076/Plano_ABC_VERSAO_FINAL_13jan2012.pdf

Digitized by Google

BRASIL. **Plano ABC -TOCANTINS: Estratégia de Elaboração. Reunião Nacional do Plano ABC**. 2014. Disponível em:<http://www.agricultura.gov.br/desenvolvimento-sustentavel/plano-abc/eventos-reunioes >

BRACELPA, Associação Brasileira de Celulose e Papel. **The 4FS Challenge Spotlight on Brazil Panel**. 2013. Disponível em <http://www.pwc.com/ca/en/forest-paper-packaging/publications/pwc-elizabeth-de-carvalhaes-presentation-2013-05-en.pdf>. Acessado em 20/01/2014.

CHEN, C. S.; TARDIN, A. T.; BATISTA, G.T. **Índices de Vegetação e suas aplicações na agricultura.** Instituto Nacional de Pesquisas Espaciais, INPE, São José dos Campos, São Paulo, 1986. 24p.

CHANDER, G., & MARKHAM, B. (2003). **Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges**. *Geoscience and Remote Sensing, IEEE Transactions on*, *41*(11), 2674-2677.

CIPRIANI, H. N. **Silvicultura para a Amazônia.** Embrapa Rondônia. Disponível em < http://www.cpafro.embrapa.br/noticia/303/?enviar_por_email=True>. Acessado em (21/10/2013).

BRITO, José Otávio. **O uso energético da madeira**. Estud. av. [online]. 2007, vol.21, n.59, pp. 185-193. ISSN 0103-4014. http://dx.doi.org/10.1590/S0103-40142007000100015.

CELINSKI, Tatiana Montes. **Classificação de cobertura do solo utilizando árvores de decisão e sensoriamento remoto**. 2008. Tese de Doutorado. UNIVERSIDADE ESTADUAL PAULISTA.

COHEN, J. **A coefficient of agreement for nominal scales**. Educational and Psychological Measurement.v.20, 1960, p.37-46.

COUTO, L.; MULLER, M. D. **Florestas energéticas no Brasil**. In: CORTEZ, L. A. B.; LORA, E. E. S.; GÓMEZ, E. O. (Ed.). Biomassa para energia. Campinas: Unicamp, 2008.

COUTO, L.;FONSECA, E,M.B.; ,MÜLLER, M.D. O **estado da arte das plantações de florestas de rápido crescimento para produção de biomassa para energia em Minas Gerais: Aspectos Técnicos, Econômicos Sociais e Ambientais**. Belo Horizonte - MG: CEMIG, 2000. 44p.

Digitized by Google

CRÓSTA, A. P. **Processamento Digital de Imagens de Sensoriamento Remoto**. Universidade Estadual de Campinas, Campinas, 4 ed. 2002.

DIAS-FILHO, M. B. **Sistemas silvipastoris na recuperação de pastagens tropicais degradadas**. Gonzaga Neto, S.; Costa, RG; Pimenta Filho, EC, p. 535-553, 2006.

DIAS-FILHO, M. B. "**Os desafios da produção animal em pastagens na fronteira agrícola brasileira**." Revista Brasileira de Zootecnia 40 (2011): 243-252.

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).

EMBRAPA: **ESTADO-DA-ARTE DA AGRICULTURA E PECUÁRIA DO TOCANTINS**. 2012. Disponível em: http://cnpasa.sede.embrapa.br/downs/estadoagropecuariato.pdf.

EMBRAPA: **Tecnologia sustentável: satélites ajudam a identificar áreas degradadas**. Disponível em <https://www.embrapa.br/busca-de-noticias/-/noticia/1478370/rio20satelites-ajudam-a-identificar-areas-degradadas>. Acesso em: 04 de julho 2014.

FAO. **Global Forest Products Facts and Figures**. 2012. Disponível em: <http://www.fao.org/forestry/statistics/80938/en/>. Acesso em: 10 jan. 2014.

FAO. **Forest Resources Assement**. 2011. Disponível em: <http://www.fao.org/forestry/fra>. Acesso em: 10 jan. 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, 1996a.

FAYYAD, U.; DJORGOVSKI, G.; WEIR, N. **Automating the Analysis and Cataloging of Sky Surveys**. In Advances in Knowledge Discovery and Data Mining, 3 ed, p. 471-493, 1996b.

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/e7a8d3c90f9f822406 da324d7ec9089d>. Acesso em:16.11.2013.

Digitized by Google

FLORENZANO, T. G. **Imagens de satélites para estudos ambientais**. In NASCIMENTO, H. R. O uso das geotecnologias para o monitoramento da agricultura de energia: Pedro Afonso (TO). Dissertação de Mestrado - Universidade Federal do Tocantins. Palmas. 2011.

FOELKEL, C. E. B. **A energia pode ser renovável, mas o modelo é sustentável?** Revista Opiniões. Disponível em: http://www.revistaopinioes.com.br/cp/materia.php?id=779. Acesso em: 15 de novembro de 2013.

FONTES, A. A.A **cadeia produtiva da madeira para energia.** 134 f. Tese (Doutorado em Ciências Florestais) - Universidade Federal de Viçosa, Viçosa, 2005. STCP. **Competitividade no Setor Florestal.** 2012.

GOLDEMBERG, J. **Pesquisa e Desenvolvimento na área de energia**. São Paulo em Perspectiva, São Paulo, v.14, n.3, p. 91 – 97, 2000. Disponível em: <http:/www.scielo.br/pdf/ssp/v14n3/9776.pdf.> Acesso em: 13/11/2013.

GOLDENBERG, J.; LUCON O. **Energia e Meio Ambiente no Brasil.** Revista Estudos Avançados, vol 21, N59, São Paulo, 2007.

GRAUER, A. KAWANO, M. **Uso de Biomassa para Produção de Energia**. Disponível em <www.ambientebrasil.com.br>. Acesso em 10/11/2013.

HALL, F. G., Townshend, J. R., Engman, E. T., **Status of remote sensing algorithms for estimation of land surface state parameters**, Remote Sensing of Environnment,51:138-156, 1995.

GRAHAM, R. L.; ENGLISH, B. C.; NOON, C. E. A Geographic Information System-based modeling system for evaluating the cost of delivered energy crop feedstock. Biomass and Bioenergy, v. 18, n. 4, p. 309-329, 2000. ISSN 09619534. Disponível em: <http://www.sciencedirect.com/science/article/B6V223YS38CH-6/2/577e4430101b28329cda80d8a001fff2 >. Acesso em: 05.04.2013.

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.

Digitized by Google

JORGE, L. G. Pereira1. Getulio T. Batista1. Dar Roberts2. **Reflectância de Coberturas Vegetais na Amazônia**. INPE - Instituto Nacional de Pesquisas Espaciais. 1996. Disponível em < http://marte.sid.inpe.br/col/sid.inpe.br/deise/1999/01.29.11.41/doc/T262.pdf>. Acessado em 20/12/2013.

JUVENAL, T.L.; MATTOS, R.L.G. **O Setor Florestal no Brasil e a Importância do Reflorestamento**. BNDS Setorial, Rio de Janeiro, n. 16, p. 3-30, set. 2002.

LANDIS, J.R.; KICH, G.G. **The measurement of observer agreement for categorical data. Journal of Biometrics**. Michigan, vol. 33, p.159-174, 1977.

LEITE, N.B. **As florestas plantadas: governo e empresas na direção do desenvolvimento sustentável**. Disponível em: http://www.sbs.org.br/secure/palestra-download.php. Acesso em: 15 de novembro de 2012.

LEITE, N.B. **Novas fronteiras: desafios e oportunidades**. Revista Opiniões. Disponível em: http://www.revistaopinioes.com.br/cp/materia.php?id=737. Acesso em: 15 de novembro de 2012.

LENZ, A. R. **Utilizando técnicas de aprendizado de máquina para apoiar o teste de regressão.** Dissertação de Mestrado. Universidade Federal do Paraná (UFPR). 2009

LILLESAND, T. M.; KIEFER, R. W. **Remote sensing and image interpretation**. 3. ed. United States of América: John Wiley & Sons, 1994. 1040 p. (ISBN 8535211772).

MACEDO, M. C. M.; Kichel, A. N.; ZIMMER, A. H. **Degradação e alternativas de recuperação e renovação de pastagens**. Campo Grande: EMBRAPA- CNPGC, 2000. 4 p. (Comunicado Técnico, 62).

MANNILA, H. Data **mining: machine learning, statistics and databases**. International Conference on Statistics and Scientific Database anagement, Estocolmo, 8, 1996.

MARTINS, F. R.; PEREIRA, E. B.; YAMASHITA, C.; PEREIRA, S. V.; MANTELLI. **MDA MATERIAL DIDÁTICO DESENVOLVIMENTO SUSTENTÁVEL.** Disponível em: http://www.bancopire.org/teceroutromundo/admin/download/DESENVOLVIMENTO%20SUSTENTAVEL.pdf.

Digitized by Google

MEIRA, C. A. A., **Processo de descoberta de conhecimento em bancos de dados para a análise e o alerta de doenças de culturas agrícolas e sua aplicação na ferrugem do cafeeiro**. 220p. Tese (Doutorado em Engenharia Agrícola) – Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas, 2008.

MENDES, JEFFERSON B. "**Estratégias e mecanismos financeiros para florestas plantadas**". MBA em Gestão Empresarial, UFPR-FGV. Curitiba, 2005. Disponível em: http://www.fao.org/forestry/12075-06238d2267638fe1c5a6f26abaa6fb6ef.pdf, acessado em março de 2014).

MENESES, P. R.; ALMEIDA, T. (org.) **Introdução ao processamento de imagens de sensoriamento remoto**. Brasília: CNPq, 2012. 266 p.

MIURA, A. K. **Identificação de áreas potenciais para cultivos destinados à produção de bioenergia com apoio de geotecnologias**. Tese de Doutorado - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2012.

MONTEBELLO, Adriana Estela Sanjuan and BACHA, Carlos José Caetano. **Avaliação das pesquisas e inovações tecnológicas ocorridas na silvicultura e na produção industrial de celulose no Brasil**. Rev. Econ. Sociol. Rural [online]. 2009, vol.47, n.2 pp. 485-517. Available from: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-20032009000200008&lng=en&nrm=iso>.

MONTIBELLER-FILHO, G. **O Mito do Desenvolvimento Sustentável:** meio ambiente e custos sociais no moderno sistema produtor de mercadorias, 2008.

MORA, Ad. L.; GARCIA, C. H. **A Cultura do eucalipto no Brasil. Sociedade Brasileira de Silvicultura. São Paulo**. 2000.

MOREIRA, R. da C. **Influência do posicionamento e da largura de bandasde sensores remotos e dos efeitos atmosféricos na determinação de índices de vegetação**. 179p. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto de Pesquisas Espaciais – São José dos Campos, 2000.

NASCIMENTO, E. P. e VIANNA, J. N. (orgs.). **Dilemas e desafios do desenvolvimento sustentável no Brasil**. Rio de Janeiro: Garamond, 2009.

NASCIMENTO, C. R. (2006). **Correção atmosferica de imagens do sensor AVHRR/NOAA utilizando produtos atmosfericos do sensor MODIS/TERRA**. Dissertação de Mestrado. Universidade Estadual de Campinas (UNICAMP). 2006.

Digitized by Google

NASCIMENTO, H. R., & DEABREU, Y. V. (2012). **Geração de informações sobre a agricultura de energia por meio das geotecnologias**. Revista Internacional de Desenvolvimento Local, 13(2), 181.

NETO, S. L. **Base de dados climático-ambientais aplicados ao setor energético – Projeto SONDA.** Anais XIII Simpósio de Sensoriamento Remoto, Goiânia-Brasil, 16-21 abril de 2005, INPE, p. 3563-3570.

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) - Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas.

ORTEGA, E.; OLIVEIRA, R.; ALBUQUERQUE, T.; ALMEIDA, A. W. **Conceitos básicos sobre sistemas agrosilvipastoris (SASP)**. 2008. Cartilha. Laboratório de Engenharia Ecológica da Unicamp – LEIA. Disponível em: < http://www.unicamp.br/fea/ortega/extensao/modulo2.pdf>. Acessado em 20/01/2014.

Plano Nacional de Agroenergia 2006-2011. Ministério da Agricultura, Pecuária e Abastecimento, Secretaria de Produção e Agroenergia. 2. ed. rev. Brasília, DF: Embrapa Informação Tecnológica, 2006.

PONZONI, F. J.; SHIMABUKURO, Y. E. **Sensoriamento remoto no estudo da vegetação**. São José dos Campos. Editora Parentese, 2007.

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>.

REZENDE, S. O.; PUGLIESI, J. B.; MELANDA, E. A.; PAULA, M. F. Mineração de dados. In REZENDE, S.O. (Org.) **Sistemas inteligentes: fundamentos e aplicações.** Barueri: Editora Manole, 2002. p. 307-335.

RIBEIRO, J. R. D. P. **A detecção remota no inventário florestal: análise das potencialidades da utilização das imagens de satélite**. 2008.SILVA, J. C.; XAVIER, B. A. **Eucalipto: manual prático do fazendeiro florestal, produzindo madeira com qualidade.** Viçosa, MG. 2006.

SANTOS, P. E. T. **Florestas Energéticas**. Revista Opiniões. Dez 2011-Fev 2012. Disponível em:< http://www.revistaopinioes.com.br/cp/materia.php?id=794>. Acessado em: 10/10/2013.

Digitized by Google

SCHOWENGERDT, R. A. **Techniques for image processing and classification in remote sensing**. New York: Academic Press, 1983. 249 p. (ISBN 0126289808).

SILVA C., NASCIMENTO H. R. e ABREU Y. V., "**PANORAMA DA BIOMASSA FLORESTAL PRIMÁRIA NO BRASIL**". In: ABREU, Y. V., OLIVEIRA, M. G. e GUERRA, S. M. G., (orgs). Energia, Sociedade e Meio Ambiente. Málaga: Eumed.net, 2010 (Disponível em: www.eumed.net/libros/2010c/723/ , acessado em 13/07/2012).

SILVA, E. P. **Fontes renováveis de energia para o desenvolvimento sustentável**. ComCiência, n. 61, 2004. Disponível em:<http://www.comciencia.br/reportagens/2004/ 12/15.shtml>. Acesso em: 15 nov. 2013.

SILVA, M. P. **Mineração de Dados-Conceitos, Aplicações e Experimentos com Weka**. (2004).

SFB. **As florestas plantadas**. Serviço Nacional de Informações Florestais – SNIF. 2013. Disponível em http://www.florestal.gov.br/snif/recursos-florestais/as-florestas-plantadas. Acessado em: 10/10/2013.

SLATER, P.N. **Remote sensing: optics and optical systems**. Reading, MA, Addison-Wesley, 1980. 575p.

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**, 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 Abril de 2014.

VIDAL, A. C. F; HORA, A. B. 2011. **Perspectivas do setor de biomassa de madeira para a geração de energia**. BNDS. Disponível em < http://www.bndes.gov.br/SiteBNDES/bndes/bndes_pt/Institucional/Publicacoes/Consulta_Expressa/Setor/Energia/201103_08.html>. Acessado em 20/10/2013.

Digitized by Google

VILELLA L.; **Recuperação de pastagens degradadas**: Seminário de Sensibilização e Difusão do Programa ABC – Estado do Tocantins. Palmas 2010.

ZULLO JUNIOR, Jurandir. **Correção atmosférica de imagens de satélite e aplicações.** 189p. Tese (Doutorado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica, Universidade Estadual de Campinas, Campinas, 1994.

WITTEN, I. H.; FRANK, E. **Data mining: pratical machine learning tools and techniques**. 2nd ed. San Francisco: Morgan Kaufmann, 2005. 525 p

Digitized by Google

APÊNDICE I - Rotina de classificação do modelo de melhor acurácia

```
=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 100

Relation: CarlosFinalDataMart2-weka.filters.unsupervised.attribute.Remove-R1,
weka.filters.unsupervised.attribute.Remove-R8-60

Instances:    52847
Attributes:   7
              Atributo
              B7
              B5
              B4
              B3
              B2
              B1
Test mode:split 66.0% train, remainder test

=== Classifier model (full training set) ===

J48 pruned tree
------------------

B7 <= 0.0774
|   B5 <= 0.0402: Corpos_dagua (8646.0)
|   B5 > 0.0402
|   |   B2 <= 0.0564
|   |   |   B3 <= 0.0398
|   |   |   |   B5 <= 0.0776
|   |   |   |   |   B4 <= 0.2661: Vegetacao (112.0/12.0)
|   |   |   |   |   B4 > 0.2661: Refloresta_1 (171.0/25.0)
|   |   |   |   B5 > 0.0776: Vegetacao (362.0/33.0)
|   |   |   B3 > 0.0398: Refloresta_1 (465.0/30.0)
|   |   B2 > 0.0564
|   |   |   B5 <= 0.1643
|   |   |   |   B3 <= 0.0423: Vegetacao (2369.0/69.0)
|   |   |   |   B3 > 0.0423
|   |   |   |   |   B2 <= 0.0597: Refloresta_1 (292.0/25.0)
|   |   |   |   |   B2 > 0.0597
|   |   |   |   |   |   B3 <= 0.0449: Vegetacao (238.0/19.0)
|   |   |   |   |   |   B3 > 0.0449
|   |   |   |   |   |   |   B2 <= 0.0697: Refloresta_1 (271.0/42.0)
|   |   |   |   |   |   |   B2 > 0.0697: Vegetacao (2172.0/35.0)
|   |   |   B5 > 0.1643: Refloresta_1 (199.0/26.0)
B7 > 0.0774
|   B7 <= 0.1818
|   |   B1 <= 0.1169
|   |   |   B4 <= 0.2242
|   |   |   |   B5 <= 0.1856
|   |   |   |   |   B5 <= 0.1734: Area_Urb (484.0/50.0)
|   |   |   |   |   B5 > 0.1734
|   |   |   |   |   |   B3 <= 0.1085: Solo_exposto (253.0/72.0)
|   |   |   |   |   |   B3 > 0.1085: Area_Urb (100.0/1.0)
|   |   |   |   B5 > 0.1856
|   |   |   |   |   B1 <= 0.0983
|   |   |   |   |   |   B4 <= 0.1902: Solo_exposto (697.0/10.0)
|   |   |   |   |   |   B4 > 0.1902
|   |   |   |   |   |   |   B7 <= 0.1586: Solo_exposto (421.0/93.0)
|   |   |   |   |   |   |   B7 > 0.1586: Refloresta_1 (115.0/35.0)
|   |   |   |   |   B1 > 0.0983
|   |   |   |   |   |   B1 <= 0.1095
|   |   |   |   |   |   |   B5 <= 0.2226
|   |   |   |   |   |   |   |   B5 <= 0.1957: Solo_exposto (162.0/60.0)
|   |   |   |   |   |   |   |   B5 > 0.1957
|   |   |   |   |   |   |   |   |   B4 <= 0.1933: Solo_exposto (100.0/27.0)
|   |   |   |   |   |   |   |   |   B4 > 0.1933: Refloresta_1 (412.0/71.0)
|   |   |   |   |   |   |   B5 > 0.2226: Refloresta_1 (662.0/11.0)
|   |   |   |   |   |   B1 > 0.1095: Area_Urb (183.0/57.0)
|   |   |   B4 > 0.2242
|   |   |   |   B5 <= 0.1961
```


Digitized by Google

```
|   |   |   |   |   B7 <= 0.1035
|   |   |   |   |   |   B4 <= 0.2452: Vegetacao (117.0/39.0)
|   |   |   |   |   |   B4 > 0.2452: Refloresta_1 (610.0/28.0)
|   |   |   |   |   B7 > 0.1035: Area_Urb (211.0/67.0)
|   |   |   |   B5 > 0.1961
|   |   |   |   |   B1 <= 0.1112: Refloresta_1 (10354.0/180.0)
|   |   |   |   |   B1 > 0.1112
|   |   |   |   |   |   B5 <= 0.2441: Area_Urb (187.0/28.0)
|   |   |   |   |   |   B5 > 0.2441: Refloresta_1 (1546.0/69.0)
|   |   B1 > 0.1169
|   |   |   B5 <= 0.2526: Area_Urb (1052.0/13.0)
|   |   |   B5 > 0.2526
|   |   |   |   B4 <= 0.247: Solo_exposto (196.0/38.0)
|   |   |   |   B4 > 0.247: Refloresta_1 (242.0/48.0)
|   B7 > 0.1818
|   |   B5 <= 0.3004
|   |   |   B7 <= 0.2137
|   |   |   |   B5 <= 0.2744
|   |   |   |   |   B3 <= 0.1169
|   |   |   |   |   |   B4 <= 0.2033
|   |   |   |   |   |   |   B5 <= 0.2158: Area_Urb (120.0/59.0)
|   |   |   |   |   |   |   B5 > 0.2158: Solo_exposto (108.0/9.0)
|   |   |   |   |   |   B4 > 0.2033: Area_Urb (120.0/43.0)
|   |   |   |   |   B3 > 0.1169
|   |   |   |   |   |   B5 <= 0.2632: Area_Urb (3001.0/35.0)
|   |   |   |   |   |   B5 > 0.2632
|   |   |   |   |   |   |   B7 <= 0.1934: Solo_exposto (180.0/103.0)
|   |   |   |   |   |   |   B7 > 0.1934: Area_Urb (421.0/13.0)
|   |   |   |   B5 > 0.2744
|   |   |   |   |   B3 <= 0.1559
|   |   |   |   |   |   B4 <= 0.2417
|   |   |   |   |   |   |   B5 <= 0.2933: Area_Urb (126.0/78.0)
|   |   |   |   |   |   |   B5 > 0.2933: Solo_exposto (116.0/4.0)
|   |   |   |   |   |   B4 > 0.2417: Refloresta_1 (492.0/110.0)
|   |   |   |   |   B3 > 0.1559
|   |   |   |   |   |   B7 <= 0.1934: Solo_exposto (194.0/2.0)
|   |   |   |   |   |   B7 > 0.1934
|   |   |   |   |   |   |   B5 <= 0.2825: Area_Urb (110.0/45.0)
|   |   |   |   |   |   |   B5 > 0.2825: Solo_exposto (310.0/44.0)
|   |   |   B7 > 0.2137: Area_Urb (5530.0/80.0)
|   |   B5 > 0.3004
|   |   |   B7 <= 0.2377
|   |   |   |   B4 <= 0.2661
|   |   |   |   |   B1 <= 0.119: Solo_exposto (2541.0/50.0)
|   |   |   |   |   B1 > 0.119
|   |   |   |   |   |   B7 <= 0.2282: Solo_exposto (653.0/57.0)
|   |   |   |   |   |   B7 > 0.2282
|   |   |   |   |   |   |   B5 <= 0.3188: Area_Urb (132.0/52.0)
|   |   |   |   |   |   |   B5 > 0.3188: Solo_exposto (152.0/14.0)
|   |   |   |   B4 > 0.2661
|   |   |   |   |   B1 <= 0.1212: Refloresta_1 (132.0/55.0)
|   |   |   |   |   B1 > 0.1212: Solo_exposto (193.0/49.0)
|   |   |   B7 > 0.2377
|   |   |   |   B5 <= 0.3248
|   |   |   |   |   B7 <= 0.2485
|   |   |   |   |   |   B5 <= 0.3127: Area_Urb (307.0/22.0)
|   |   |   |   |   |   B5 > 0.3127: Solo_exposto (210.0/77.0)
|   |   |   |   |   B7 > 0.2485: Area_Urb (1156.0/20.0)
|   |   |   |   B5 > 0.3248
|   |   |   |   |   B3 <= 0.1783: Solo_exposto (2208.0/171.0)
|   |   |   |   |   B3 > 0.1783
|   |   |   |   |   |   B4 <= 0.2801: Area_Urb (473.0/73.0)
|   |   |   |   |   |   B4 > 0.2801
|   |   |   |   |   |   |   B3 <= 0.1954: Solo_exposto (312.0/34.0)
|   |   |   |   |   |   |   B3 > 0.1954: Area_Urb (149.0/31.0)

Number of Leaves  :     57

Size of the tree :      113


Time taken to build model: 1.83 seconds
```


Digitized by Google

```
=== Evaluation on test split ===
=== Summary ===

Correctly Classified Instances       16857               93.8168 %
Incorrectly Classified Instances      1111                6.1832 %
Kappa statistic                          0.92
Mean absolute error                      0.0381
Root mean squared error                  0.1411
Relative absolute error                 12.3125 %
Root relative squared error             35.8713 %
Total Number of Instances            17968

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
0.907     0.017      0.95      0.907     0.928      0.986    Area_Urb
0.939     0.005      0.957     0.939     0.948      0.998    Vegetacao
0.957     0.031      0.931     0.957     0.944      0.99     Refloresta_1
0.895     0.028      0.868     0.895     0.881      0.981    Solo_exposto
1         0          1         1         1          1        Corpos_dagua
Weighted Avg.    0.938     0.019      0.939     0.938     0.938      0.99

=== Confusion Matrix ===

    a    b    c    d    e   <-- classified as
 4349    1  116  328    0 |    a = Area_Urb
    6 1730  106    0    0 |    b = Vegetacao
   63   77 5182   91    0 |    c = Refloresta_1
  160    0  162 2756    0 |    d = Solo_exposto
    1    0    0    0 2840 |    e = Corpos_dagua
```


Digitized by Google

ANEXO I - Lista de imóveis com reflorestamentos por município no Estado do Tocantins, coordenadas geográficas e a área utilizada em cada propriedade

Regional	Município	Propriedade	Proprietário	Latitude	Longitude	Área	Espécie/Variedade
Araguaina	Araguaina	Faz. Santiago	Ecobrasil			1.148,93	Eucalipto
Araguaina	Babaçulandia	Faz. Alvorada	Ecobrasil			4.139,20	Eucalipto
Araguaina	Wanderlância	Faz. Clarão da Lua	Ecobrasil			4.569,00	Eucalipto
Araguaina	Wanderlância	Faz. Monte Cristo	Ecobrasil			1.393,37	Eucalipto
Araguaina	Xambioá	Faz. Beira Rio	José Antônio Pereira	06 35 16,26	48 21 26	200,00	Eucalipto
Araguaina	Piraquê	Faz. Batateira	Leandro Bernardino Pereira	06 46 43	48 17 49,48	43,00	Eucalipto
Araguaina	Santa Fé	Faz. Novo Horizonte	Eurides Facchine	07 05 59	48 45 33,4	6,00	Eucalipto
Araguaina	Santa Fé	Faz. Novo Horizonte	Eurides Facchine	07 05 59	48 45 33,4	505,00	Seringueira/RIM600
Araguaina	Babaçulandia	Faz. Santa Isabel	Ana Cecília M. Andrade			20,00	Eucalipto
Araguatins	Araguatins	Faz. Tobasa	Suzano Papel e Celulose	05 51 40	47 53 30	287,36	Eucalipto
Araguatins	Araguatins	Faz. Araguari	Suzano Papel e Celulose	05 46 54	47 56 53	254,34	Eucalipto
Araguatins	Araguatins	Faz. São Salvador I	Sinobras			320,77	Eucalipto
Araguatins	Lagoa de São Bento	Faz. São Salvador II	Sinobras	05 59 16,5	48 06 28,7	417,56	Eucalipto
Araguatins	Lagoa de São Bento	Faz. São Martinho	Sinobras	05 56 41,7	48 02 38,3	2.559,25	Eucalipto
Araguatins	Lagoa de São Bento	Faz. Brejo Grande	Sinobras	05 57 59,6	48 02 47,3	2.676,27	Eucalipto
Araguatins	Lagoa de São Bento	Faz. Ronca	Sinobras	06 00 50	48 02 22,3	983,82	Eucalipto
Araguatins	Lagoa de São Bento	Faz. Santa Maria	Sinobras	06 00 44,3	48 04 46,8	297,12	Eucalipto
Marabá	Lagoa de São Bento	Faz. Santa Lucia	Sinobras	06 02 32,4	48 02 5,7	2.831,93	Eucalipto
Araguatins	Lagoa de São Bento	Faz. São Francisco	Sinobras			634,84	Eucalipto
Araguatins	Araguatins	Faz. Novo Cerrado	Sinobras			300,00	Eucalipto
Araguatins	Araguatins	Faz. Estiva	Sinobras	05 45 12,5	48 08 00,1	2.336,00	Eucalipto
Araguatins	Lagoa de São Bento	Faz. Santa Lucia II	Sinobras			90,00	Eucalipto
Araguatins	Lagoa de São Bento	Faz. Campos Belos	Erípedes Antonio Vieira	05 34 52,7	48 03 26,6	40,00	Eucalipto
Araguatins	São Miguel	Nobleinvest Ativ. Rurais	Nobleinvest Ativ. Rurais LTDA	05 32 35,8	47 33 18,1	68,30	Acácia
Araguatins	São Miguel	Nobleinvest Ativ. Rurais	Nobleinvest Ativ. Rurais LTDA	05 32 35,8	47 33 18,1	158,00	Gmelina
Araguatins	São Miguel	Nobleinvest Ativ. Rurais	Nobleinvest Ativ. Rurais LTDA	05 32 35,8	47 33 18,1	44,00	Mogno
Araguatins	São Miguel	Nobleinvest Ativ. Rurais	Nobleinvest Ativ. Rurais LTDA	05 32 35,8	47 33 18,1	19,00	Paricá
Araguatins	São Miguel	Nobleinvest Ativ. Rurais	Nobleinvest Ativ. Rurais LTDA	05 32 35,8	47 33 18,1	1.223,90	Teca
Colinas	Couto Magalhães	Faz. Abacaxizal (Faz. C	Ecobrasil			60 ou 94	Eucalipto
Colinas	Juarina	Faz. Nova Fronteira	Ecobrasil			135,52	Acácia
Colinas	Palmeirante	Faz. Bom Jesus	Ecobrasil			380,33	Eucalipto
Colinas	Palmeirante	Faz. São Judas Tadel	Ecobrasil			213,64	Eucalipto
Colinas	Palmeirante	Faz. Bananal	Ecobrasil			423,42	Eucalipto
Colinas	Palmeirante	Faz. Prata	Ecobrasil			2.392,74	Eucalipto

Digitized by Google

Colinas	Palmeirante	Faz. Riachinho	Ecobrasil			163,15	Eucalipto
Colinas	Palmeirante	Faz. Mangabal	Ecobrasil			98,88	Eucalipto
Colinas	Palmeirante	Faz. Taboca	Ecobrasil			336,15	Eucalipto
Colinas	Palmeirante	Faz. São Paulo	Ecobrasil			1.179,81	Eucalipto
Colinas	Palmeirante	Faz. Carretão	Ecobrasil			46,07	Eucalipto
Pedro Afonso	Tupiratins	Faz. Paraíso	Vilmar Zuki			100,00	Eucalipto
Pedro Afonso	Tupiratins	Faz. Santo Expedito	Percival Peluso			65,00	Eucalipto
Pedro Afonso	Tupiratins	Faz. Três meninos	Antonio Gonzaga			200,00	Eucalipto
Pedro Afonso	Tupiratins	Faz. Nº Sª Aparecida				300,00	Eucalipto
Pedro Afonso	Itapiratins	Faz. Belo Monte	Eterro Flávio Ricardo				Eucalipto
Pedro Afonso	Itacajá	Faz. Riva	Emani Riva			70,00	Eucalipto
Pedro Afonso	Itacajá	Faz. Maman	Namir Antonio Maman			30,00	Eucalipto
Pedro Afonso	Pedro Afonso	Faz. Brejinho				150,00	Seringueira
Pedro Afonso	Goiatins		Ecobrasil				Eucalipto
Pedro Afonso	Barra do Ouro		Ecobrasil				
Pedro Afonso	Centenário	Faz. São Sebastião	Alvino Moteiro Dias				Teca
Pedro Afonso	Centenário	Faz. Estoquimas					Eucalipto
Pedro Afonso	Centenário	Faz. Aciole Sales					Eucalipto
Pedro Afonso	Centenário	Faz. Nossa	Geraldo Cancelier			200	Eucalipto
Pedro Afonso	Centenário	Faz. Maça Verde	Mário Back			100	Eucalipto
Pedro Afonso	Centenário	Faz. Moeda					Eucalipto
Pedro Afonso	Centenário	Faz. Rio Norte	Wilson Rezende Alexander			400	Eucalipto
Pedro Afonso	Centenário	Faz. Sagrada Família	Eder			500	Eucalipto
Pedro Afonso	Centenário	Faz. Boa Esperança I	Edivar Garcia de Paula			387	Eucalipto
Pedro Afonso	Centenário	Faz. Brejo Grande	Afonso Celso Silveira			96,8	Eucalipto
Pedro Afonso	Centenário	Faz. Bela Vista				120	Eucalipto
Pedro Afonso	Campos Lindos		Bungue Alimentos				Eucalipto
Pedro Afonso	Campos Lindos	Faz. Taboquinha	Fernando Lira de Abreu				Eucalipto
Pedro Afonso	Goiatins	Faz. São João	Ecobrasil			895,37	Eucalipto
Pedro Afonso	Goiatins	Faz. Talismã	Ecobrasil			1580,24	Eucalipto
Pedro Afonso	Goiatins	Faz. Boa Vista	Ecobrasil			129,03	Eucalipto
Pedro Afonso	Goiatins	Faz. Ribeiraozinho	Ecobrasil			154,66	Eucalipto
Pedro Afonso	Goiatins	Faz. Quebrada	Ecobrasil			577,08	Eucalipto
Pedro Afonso	Barra do Ouro	Faz. Ilha do Porto	Ecobrasil			1464,12	Eucalipto
Pedro Afonso	Barra do Ouro	Faz. Água Azul	Ecobrasil			1368,36	Eucalipto
Pedro Afonso	Barra do Ouro	Faz. Tauã	Ecobrasil			2993,5	Eucalipto
Pedro Afonso	Barra do Ouro	Faz. Altamira	Ecobrasil			1057,48	Eucalipto
Pedro Afonso	Centenário						
Miracema	Dois irmãos	Faz. Esmeralda	João Benjamim Gomes	09 31 46	48 53 51	200	Eucalipto

Digitized by Google

Miracema	Dois irmãos	Faz. Três poderes	Jorge Alves Toledo	09 28 58	48 51 58	200	Eucalipto
Miracema	Dois irmãos	Faz. Três poderes	Jorge Alves Toledo	09 28 58	48 51 58	5	Seringueira
Miracema	Dois irmãos	Faz. Palmeirinha	Leandro Gasparine	09 38 29	49 00 79	134	Eucalipto
Miracema	Miracema	Faz. Terra Nossa	Antonio Paulino Alves	09 32 15	48 32 55	53	Eucalipto
Miracema	Miracema	Faz. Santa Vitória	Cinésio Barbosa	09 36 07,2	48 35 09,4	25	Teca
Miracema	Miracema	Faz. Santa Lourdes	Jeremias Garcia Soares	09 31 45,7	48 27 26,1	242	Eucalipto
Miracema	Miracema	Faz. Brandão	Joaquim Vieira dos Santos	09 43 36,1	48 46 00,5	15	Eucalipto
Miracema	Miracema	Faz. Xinguara	Rosano Tondato	09 42 39,8	48 22 58,1	60	Eucalipto
Miracema	Miranorte	Faz. Agua Limpa	Dionilson Mendes	09 25 04,9	48 34 17,5	22	Eucalipto
Miracema	Miranorte	Faz. Tapejara	Eronita Bressan	09 31 56,8	48 49 21,8	125	Eucalipto
Miracema	Miranorte	Faz. Pequizeiro	Valdez Coelho Carvalho	09 33 02,4	48 50 26,6	18,5	Eucalipto
Miracema	Miranorte	Faz. Bacaba	Leandro Gasparine	09 31 60	48 48 34,1	100	Eucalipto
Paraiso	Pugmil	Faz. São João	Cotril Floresta			900	Eucalipto
Paraiso	Pugmil	Faz. Cabeceira Verde	Manoel Silva Nunes			5	Seringueira
Paraiso	Pugmil	Faz. Sítio Atlantis	Bonifácio			19,8	Seringueira
Porto Nacional	Brejinho de Nazaré	Faz. Angelina	John George de Carle			138	Neen
Porto Nacional	Brejinho de Nazaré	Faz. Angelina	John George de Carle			12	Eucalipto
Porto Nacional	Brejinho de Nazaré	Faz. Angelina	John George de Carle			1	Paricá
Porto Nacional	Brejinho de Nazaré	Faz. Angelina	John George de Carle			9	Pequi
Porto Nacional	Brejinho de Nazaré	Faz. Angelina	John George de Carle			80	Teca
Porto Nacional	Silvanópolis	Faz. Bom Sucesso	Armiro Abentroth			25	Eucalipto
Porto Nacional	Monte do Carmo	Faz. Serra Geral	Antonio Carlos V. Golveia				Eucalipto
Porto Nacional	Monte do Carmo	Faz. Areinha	Maria Valéria Alves			3	Eucalipto
Porto Nacional	Monte do Carmo	Faz. Boa Vista	Rodrigo Matos Lourenço				Eucalipto
Porto Nacional	Natividade	Faz. Bom Tempo	Paulo Antonio Rosa			90	Acácia
Porto Nacional	Chapada da Natividade	Faz. Prova	Carlos Gomes da Paixão			68	Seringueira
Palmas	Novo Acordo	Faz. Galega	Caio Fernande Marques	10 15 54,8	47 08 24,8	30	Seringueira
Palmas	Novo Acordo	Faz. Galega	Caio Fernande Marques	10 16 10,5	47 08 35,5	20	Eucalipto
Palmas	Novo Acordo	Faz. Galega	Caio Fernande Marques	10 16 10,5	47 08 35,5	15	Teca
Palmas	Novo Acordo	Águas de Março	João Moreira da Costa	10 08 41	47 22 09,3	2	Eucalipto
Palmas	Novo Acordo	Faz. Boa Esperança	José Leonardo Teixeira Gome	09 59 15,7	46 56 00	154	Eucalipto
Palmas	Novo Acordo	Faz. Desafio I	Marcelo Cavalcante Bazr	16 08 01	47 10 12,4	600	Eucalipto
Palmas	São Félix do Tocantins	Faz. Bela Vista	Oscar Paulo Hogemboom	10 06 47,2	47 00 41,3	300	Eucalipto
Palmas	São Félix do Tocantins	Faz. Sumidouro	Diniz Novelo e Paulo Lacerda	10 04 47,7	46 46 14,6	268	Eucalipto
Palmas	Novo Acordo	Faz. Diamante	Luis Carlos S. Carvalho	10 01 12,5	47 46 47,9	130	Eucalipto
Gurupi	Palmeiropolis	Faz. Libra	Plínio Flávio Arcanjo			120	Seringueira
Gurupi	Palmeiropolis	Faz. Serra Dourada	George Hajjar			700	Seringueira
Gurupi	Cariri	Faz. Santa Helena	José Louri Melo Barreto	11 47 29	49 07 47,2	25	Seringueira
Formoso	Formoso do Araguaia	Faz. São Carlos Araguai	Waldomiro Paccola			600	Eucalipto

Digitized by Google

Formoso	Dueré	Faz. Reunidas	Paulo Arantes Ferraz			3500	Eucalipto
Formoso	Formoso do Araguaia	Faz. São Pedro	Madeireira Jequitibá			39	Eucalipto
Formoso	Figueirópolis	Faz. Esperança	Thiago de Abreu Pulice			460	Eucalipto
Formoso	Formoso do Araguaia	Faz. Nª Sª Aparecida	Haroldo Alves Resende	11 53 16,55	49 23 48,26	12	Mogno
Taguatinga	Almas	Faz. Baixão Verde	José Américo França Vieira			200	Pinus
Taguatinga	Almas	Faz. Pontal da Serra	Gilberto Donizete Matias			150	Pinus
Taguatinga	Conceição do Tocantins	Faz. Santa Rosa	GMR Florestal			432,67	Eucalipto
Taguatinga	Conceição do Tocantins	Faz. Monte Alegre	GMR Florestal			195,21	Eucalipto
Taguatinga	Conceição do Tocantins	Faz. Caiçara	GMR Florestal			332,87	Eucalipto
Taguatinga	Conceição do Tocantins	Faz. Alvorada	GMR Florestal			555,25	Eucalipto
Taguatinga	Conceição do Tocantins	Faz. Belo Horizonte	GMR Florestal			61,09	Eucalipto
Taguatinga	Conceição do Tocantins	Faz. Posse	GMR Florestal			207,7	Eucalipto
Taguatinga	Conceição do Tocantins	Faz. Perseverança	Antonio Arrugiero Breda			400	Teca
Taguatinga	Dianópolis	Faz. Morada Verde	Cláudio Norberto Martin			120	Eucalipto
Taguatinga	Dianópolis	Faz. Thara	Rafael de Leon			500	Eucalipto
Taguatinga	Dianópolis	Faz. Itapiranga	Miguel Angelo Muzeti			270	Eucalipto
Taguatinga	Dianópolis	Faz. Orinte	Ezequiel			122	Eucalipto
Taguatinga	Novo Jardim	Faz. Entre Rios	Eduardo Manzoti			120	Eucalipto
Taguatinga	Paranã	Faz. Retiro					Eucalipto
Taguatinga	Porto Alegre do Tocantins	Faz. Bajé	São Francisco Agribussines LTDA			7	Teca
Taguatinga	Porto Alegre do Tocantins	Faz. Bajé	São Francisco Agribussines LTDA			57	Eucalipto
Taguatinga	Ponte Alta do Tocantins	Faz. Monte Cristo	Paulo Fernando Garcia Gonçalves			145	Eucalipto
Taguatinga	Ponte Alta do Tocantins	Faz. Joaquim Lote1	Vilmar Falchi			450	Eucalipto
Taguatinga	Ponte Alta do Tocantins	Faz. Joaquim Lote3	Sidnei Falchi			67	Eucalipto
Taguatinga	Ponte Alta do Tocantins	Faz. Ribeirão	Paulo Corazi			80	Eucalipto
Taguatinga	Rio da Conceição	Faz. Rancho Novo	Latex Serra Geral			41	seringueira
Taguatinga	Rio da Conceição		Rodrigo Schumacher				
Taguatinga	Taguatinga	Faz. Menina Izildinha V	Célio Cicente Anatriéllo			356	Eucalipto
Taguatinga	Taguatinga	Faz. Ema	Lozo Mário José Pereira			300	Eucalipto
Taguatinga		Faz. Bom Sucesso	Édson Lopes Ferreira			25	Eucalipto

Digitized by Google

www.lifeeditora.com.br

Adquira mais livros agora mesmo em nosso site.

Esta obra foi composta em Electra para
Life Editora em novembro de 2020.

Digitized by Google

As florestas plantadas tem atraído grande interesse pela possibilidade de utilização em aplicações bioenergéticas frente à tendência mundial de priorizar fontes de energia que proporcionem maior sustentabilidade ambiental, mais qualidade e segurança. No Brasil, os deslocamentos na geografia da cadeia produtiva agroflorestal atual em direção às regiões de fronteira agrícola (Centro-Oeste e Norte) vem criando desafios de adequação dos conhecimentos técnico-científicos já consolidados em outras regiões. Nesse contexto, o objetivo desta dissertação é avaliar a acurácia da classificação e identificação de áreas cultivadas com florestas plantadas para fins energéticos, em imagens orbitais do sensor Landsat 5 TM. Por meio de técnicas estatísticas de mineração de dados, o presente trabalho também avaliou a utilização de um amplo conjunto de atributos para identificar melhorias nos resultados da classificação. A pesquisa se concentrou em amostras de áreas plantadas no estado do Tocantins, região norte do Brasil. 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, como pela redução da quantidade de informação necessária para a resolução deste tipo de problema. Assim, as técnicas empregadas nesse estudo possibilitam o desenvolvimento de modelos de classificação robustos no auxílio ao planejamento e à tomada de decisão sobre a plantação de florestas no território brasileiro.

Estudo defendido como dissertação no Mestrado em Agroenergia da Universidade Federal do Tocantins em 2014. Os autores fizeram pequenas modificações ao transformar em livro

LIFE EDITORA

ISBN 978-65-5887-013-5

Digitized by Google