

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS RURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA FLORESTAL**

Roberta Aparecida Fantinel

**PROCEDIMENTOS DE APRENDIZAGEM DE MÁQUINA PARA ANÁLISE DE
PADRÕES ESPACIAIS COM O USO DA PLATAFORMA KNIME**

Santa Maria, RS
2020

Roberta Aparecida Fantinel

**PROCEDIMENTOS DE APRENDIZAGEM DE MÁQUINA PARA ANÁLISE DE
PADRÕES ESPACIAIS COM O USO DA PLATAFORMA KNIME**

Dissertação apresentada ao curso de Mestrado do Programa de Pós-Graduação em Engenharia Florestal, área de Concentração em Manejo Florestal, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestra em Engenharia Florestal**.

Orientador: Prof. Dr. Rudiney Soares Pereira

Santa Maria, RS
2020

Fantinel, Roberta Aparecida

Procedimentos de aprendizado de máquina para análise de padrões espaciais com o uso da plataforma KNIME / Roberta Aparecida Fantinel.- 2020.

74 p.; 30 cm

Orientador: Rudiney Soares Pereira

Coorientador: Frederico Dimas Fleig

Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós Graduação em Engenharia Florestal, RS, 2020

1. Inteligência artificial 2. KNIME 3. Landsat 4. Sensoriamento Remoto I. Soares Pereira, Rudiney II. Dimas Fleig, Frederico III. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

© 2020 Todos os direitos reservados a Roberta Aparecida Fantinel. A reprodução de parte ou do todo deste trabalho só poderá ser feita mediante a citação da fonte. Endereço eletrônico: fantinel.ar@gmail.com

Roberta Aparecida Fantinel

**PROCEDIMENTOS DE APRENDIZAGEM DE MÁQUINA PARA ANÁLISE DE
PADRÕES ESPACIAIS COM O USO DA PLATAFORMA KNIME**

Dissertação apresentada ao curso de Mestrado do Programa de Pós-Graduação em Engenharia Florestal, área de Concentração em Manejo Florestal, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestra em Engenharia Florestal**.

Aprovada em 20 de fevereiro de 2020:



Rudiney Soares Pereira, Dr. (UFSM)
(Presidente/Orientador)



Damáris Gonçalves Padilha, Dra. (UFSM)



Emanuel Araújo Silva, Dr. (UFRPE)

Santa Maria, RS
2020

AGRADECIMENTOS

Agradeço aos meus pais Vilmor e Elita Bortolás Fantinel e a minha irmã Elisângela Fantinel por me apoiarem e incentivarem em todas as minhas decisões, a ajuda foi fundamental para que eu pudesse realizar os meus sonhos. Saibam que os seus esforços não foram em vão, meu muito obrigada. Amo vocês!

Ao meu namorado Varlen Nassinhack pelo carinho, paciência e companheirismo. As que posso, na caminhada da vida, chamar de amigas: Carina Stahl, Janice Stanque, Silvana Caetano - "As garotas do jardim". As amigas e parceiras de pesquisa Fernanda Dias, Elenice Weiler, Tásilla Loiola e ao amigo André Ballestreli, obrigada pelo apoio incondicional, pelas palavras de incentivo e pelos ensinamentos. Gratidão!

A Universidade Federal de Santa Maria que, por meio do Programa de Pós-Graduação em Engenharia Florestal (PPGEF), possibilitou prosseguir meus estudos, meu desejo de que a UFSM continue pública, gratuita, de qualidade e socialmente referenciada. Aos docentes que compartilharam o seu conhecimento, contribuindo, sobremaneira, para que pudesse alcançar o objetivo final, a dissertação.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro, através da concessão de bolsa de estudos.

Ao professor Dr. Rudiney Soares Pereira pela orientação deste trabalho, minha gratidão pelas oportunidades, pelos desafios colocados e, especialmente, pela confiança, elementos essenciais que me motivam a seguir firme na caminhada científica.

Aos membros da equipe do Laboratório de Sensoriamento Remoto (LABSERE), em especial aos colegas Juliana Marchesan e Mateus Schuh, pelo acolhimento, apoio e parceria constante.

Agradeço aos membros da banca avaliadora, professora Damáris Gonçalves Padilha e ao professor Emanuel Araújo Silva, que dividiram comigo este momento tão importante e esperado.

A Deus.

A todos vocês, os meus sinceros agradecimentos.

Obrigada!

“Então é hora de recomeçar tudo outra vez, sem ilusão e sem pressa, mas com a teimosia do inseto que busca um caminho no terremoto.”

Carlos Drummond de Andrade.

RESUMO

PROCEDIMENTOS DE APRENDIZAGEM DE MÁQUINA PARA ANÁLISE DE PADRÕES ESPACIAIS COM O USO DA PLATAFORMA KNIME

AUTORA: Roberta Aparecida Fantinel
ORIENTADOR: Rudiney Soares Pereira

Imagens de satélites provenientes dos sensores remotos surgem como alternativas viáveis e eficientes no estudo de informações sobre os padrões do uso e cobertura da terra. Dados extraídos das imagens de satélite, são atualmente empregados no aprendizado de máquina, sendo este método capaz de prever a classe de novos dados do domínio no qual ele foi treinado. Desta forma, o presente trabalho teve como objetivo analisar a capacidade dos algoritmos de aprendizado de máquina em prever o uso e cobertura da terra do município de Dona Francisca - RS. O banco de dados geográficos foi implementado no *software* QGIS, onde iniciou-se a importação das imagens TM/Landsat 5 nos anos de 2004 e 2009 e OLI/Landsat 8 para os anos de 2015 e 2019. Posteriormente realizou-se a composição sintética das bandas falsa cor RGB 543 do Landsat 5 e RGB 654 Landsat 8, com a finalidade de obter as amostras dos *pixels* de referência, levando em consideração a informação espectral de cada *pixel* (valor numérico), com a finalidade de obter informações para caracterizar e diferenciar os padrões de uso e cobertura da terra (água, agricultura, campo, floresta e solo exposto). Após iniciou-se o treinamento e teste dos algoritmos nas proporções de 80%-20%, 70%-30%, 60%-40% por meio dos algoritmos de aprendizado de máquina *Random Forest* (RF), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN) e *Naive Bayes* (NB) no *software* KNIME, e por fim apresentado o desempenho da acurácia global e do índice de Kappa. Os resultados evidenciaram que os algoritmos de aprendizado de máquina RF e SVM apresentaram os melhores desempenhos para os anos de 2004 e 2009. Já para o ano de 2015, os algoritmos KNN e a RF tiveram uma acurácia global melhor. O algoritmo NB demonstrou em todos os testes desempenhos inferiores aos demais algoritmos estudados. Os valores do índice Kappa gerados pelo *software* KNIME indicam que a qualidade das classificações geradas pelos algoritmos RF, SVM, KNN e NB para todos os anos foram de muito boa a excelente. Evidencia-se que os algoritmos de aprendizado de máquina mostraram resultados satisfatórios, de maneira que apresentaram eficiência em prever o uso e cobertura da terra a partir de dados provenientes das imagens orbitais.

Palavras-chave: Inteligência artificial. Landsat. KNIME. Sensoriamento Remoto.

ABSTRACT

MACHINE LEARNING PROCEDURES FOR ANALYSIS OF SPATIAL PATTERNS WITH THE USE OF THE KNIME PLATFORM

AUTHOR: Roberta Aparecida Fantinel
ADVISOR: Rudiney Soares Pereira

Satellite images from remote sensors appear as viable and efficient alternatives in the study of information on patterns of land use and coverage. Data extracted from satellite images are currently used in machine learning, and this method is able to predict the class of new data in the domain in which it was trained. Thus, the present study aimed to analyze the capacity of machine learning algorithms to predict land use and land cover in the municipality of Dona Francisca - RS. The geographic database was implemented in the QGIS software, where the import of TM/Landsat 5 images began in 2004 and 2009 and OLI/Landsat 8 for 2015 and 2019. Subsequently, the synthetic composition of the false bands RGB color 543 from Landsat 5 and RGB 654 Landsat 8, in order to obtain the samples of the reference pixels, taking into consideration the spectral information of each pixel (numerical value), in order to obtain information to characterize and differentiate patterns of land use and coverage (water, agriculture, countryside, forest and exposed soil). After the training and testing of the algorithms started in the proportions of 80% - 20%, 70% - 30%, 60% - 40% through the machine learning algorithms Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Naive Bayes (NB) in the KNIME software, and finally presented the performance of the global accuracy and the Kappa index. The results showed that the RF and SVM machine learning algorithms showed the best performances for the years 2004 and 2009. As for the year 2015, the KNN and RF algorithms had a better overall accuracy. The NB algorithm showed lower performance in all tests than the other studied algorithms. The Kappa index values generated by the KNIME software indicate that the quality of the classifications generated by the RF, SVM, KNN and NB algorithms for all years were from very good to excellent. It is evident that the machine learning algorithms showed satisfactory results, so that they were efficient in predicting land use and land cover from data from orbital images.

Keywords: Artificial intelligence. Landsat. KNIME. Remote Sensing.

LISTA DE ILUSTRAÇÕES

Figura 1 - Processo de obtenção das imagens por meio do Sensoriamento Remoto.....	19
Figura 2 - Localização do município de Dona Francisca, Rio Grande do Sul.....	38
Figura 3 - Fluxograma dos procedimentos metodológicos.....	40
Figura 4 - Exemplo do fluxo de processos e operadores usados no modelo preditivo do algoritmo <i>Naive Bayes</i>	44
Figura 5 - Identificação das amostras dos <i>pixels</i> de referência da imagem sintética RGB 543 do sensor TM/Landsat 5, nos diferentes padrões de uso e cobertura da terra para o ano de 2004, município de Dona Francisca - RS.....	47
Figura 6 - Identificação das amostras dos <i>pixels</i> de referência da imagem sintética RGB 543 do sensor TM/Landsat 5, nos diferentes padrões de uso e cobertura da terra para o ano de 2009, município de Dona Francisca - RS.....	48
Figura 7 - Identificação das amostras dos <i>pixels</i> de referência da imagem sintética RGB 654 do sensor OLI/Landsat 8, nos diferentes padrões de uso e cobertura da terra para o ano de 2015, município de Dona Francisca - RS.....	49
Figura 8 - Identificação das amostras dos <i>pixels</i> de referência da imagem sintética RGB 654 do sensor OLI/Landsat 8, nos diferentes padrões de uso e cobertura da terra para o ano de 2019, município de Dona Francisca - RS.....	50

LISTA DE TABELAS

Tabela 1 - Características do sensor <i>Thematic Mapper</i> (TM) a bordo do satélite Landsat 5.....	22
Tabela 2 - Características do sensor <i>Operational Land Imager</i> (OLI) e <i>Thermal InfraRed Sensor</i> (TIRS) a bordo do satélite Landsat 8.....	23
Tabela 3 - Matriz de confusão.....	25
Tabela 4 - Qualidade da classificação de acordo com os intervalos do coeficiente Kappa.....	26
Tabela 5 - Imagens abrangentes do TM/Landsat 5 e OLI/Landsat 8 e as respectivas datas, Dona Francisca - RS.....	41
Tabela 6 - Acurácia global do desempenho dos algoritmos na predição dos padrões do uso e cobertura da terra, utilizando o software KNIME - outubro de 2004.....	52
Tabela 7 - Índices de Kappa obtidos pelos algoritmos de aprendizado de máquina, outubro de 2004.....	52
Tabela 8 - Acurácia global do desempenho dos algoritmos nas predições dos padrões do uso e cobertura da terra, utilizando o software KNIME - outubro de 2009.....	54
Tabela 9 - Índices de Kappa obtidos pelos algoritmos de aprendizado de máquina, outubro de 2009.....	54
Tabela 10 - Acurácia global do desempenho dos algoritmos nas predições dos padrões do uso e cobertura da terra, utilizando o software KNIME - outubro de 2015.....	55
Tabela 11 - Índices de Kappa obtidos pelos algoritmos de aprendizado de máquina, outubro de 2015.....	56
Tabela 12 - Acurácia global do desempenho dos algoritmos nas predições dos padrões do uso e cobertura da terra, utilizando o software KNIME - outubro de 2019.....	57
Tabela 13 - Índices de Kappa obtidos pelos algoritmos de aprendizado de máquina, outubro de 2019.....	58

LISTA DE QUADROS

Quadro 1 - Padrões de uso e cobertura da terra visualizada nas imagens RGB 543 do sensor TM e RGB 654 do sensor OLI.....	42
--	----

LISTA DE EQUAÇÕES

Equação 1.....	25
Equação 2.....	26
Equação 3.....	31
Equação 4.....	34

LISTA DE ABREVIATURAS

AM – Aprendizado de Máquina
CFA – Subtropical úmido
CSV – *Comma Separated Values*
ERST-1 – *Earth Resources Techological Satellite-1*
GLP – *General Public License*
GPS – Sistema de Posicionamento Global
IBGE – Instituto Brasileiro de Geografia e Estatística
KNIME – *Konstanz Information Miner*
KNN – *K-Nearest Neighbor*
LANDSAT – *Land Remote Sensing Satellite*
MSS – *Multispectral Scanner System*
NASA – *National Aeronautics & Space Administration*
NB – *Naive Bayes*
OLI – *Operational Land Imager*
OSH – *Optimal Separating Hyperplane*
PIXEL – *Picture element*
RF – *Random Forest*
SIG – Sistema de Informação Geográfica
SR – Sensoriamento Remoto
SVM – *Support Vector Machine*
TIRS – *Thermal Infrared Sensor*
USGS – *United States Geological Survey*
UTM – Universal Transversa de Mercator
WGS – *World Geodetic System*

SUMÁRIO

1 INTRODUÇÃO	16
1.1 OBJETIVOS	18
1.1.1 Objetivo Geral	18
1.1.2 Objetivos Específicos	18
2 REVISÃO DE LITERATURA	19
2.1 SENSORIAMENTO REMOTO	19
2.1.1 Satélite da série Landsat - <i>Land Remote Sensing Satellite</i>	21
2.1.2 Classificação digital e validação	24
2.2 GEOPROCESSAMENTO E SISTEMA DE INFORMAÇÃO GEOGRÁFICA	27
2.3 APRENDIZADO DE MÁQUINA	27
2.3.1 Modelos supervisionados e não supervisionados	30
2.3.2 K-Vizinhos mais próximos (<i>K-Nearest Neighbors - KNN</i>)	31
2.3.3 Máquinas de Vetores de Suporte (<i>Support Vector Machine – SVM</i>)	32
2.3.4 Florestas aleatórias (<i>Random Forest – RF</i>)	33
2.3.5 <i>Naive Bayes – NB</i>	34
2.4 SOFTWARE LIVRE PARA SENSORIAMENTO REMOTO	35
2.4.1 <i>Konstanz Information Miner – KNIME</i>	35
2.4.1 <i>Software QGIS</i>	35
3 MATERIAL E MÉTODOS	38
3.1 CARACTERIZAÇÃO DA ÁREA DE ESTUDO	38
3.1.1 Localização da área de estudo	38
3.1.2 Aspectos físicos e socioeconômicos	39
3.2 MATERIAL	40
3.3 MÉTODOS	40
3.3.1 Processamento dos dados no QGIS	41
3.3.2 Processamento das imagens	41
3.3.3 Aprendizado de máquina no <i>software KNIME</i>	43
4 RESULTADOS E DISCUSSÃO	46
4.1 PIXELS DE REFERÊNCIA	46
4.2 TREINAMENTO, TESTE E DESEMPENHO DOS ALGORITMOS	51
4.2.1 Imagem do sensor TM/Landsat 5 – outubro de 2004	51
4.2.2 Imagem do sensor TM/Landsat 5 – outubro de 2009	53

4.2.3 Imagem do sensor OLI/Landsat 8 – outubro de 2015	55
4.2.4 Imagem do sensor OLI/Landsat 8 – outubro de 2019	57
5 CONSIDERAÇÕES FINAIS	59
REFERÊNCIAS	61

1 INTRODUÇÃO

Há centenas de anos, a atividade humana vem promovendo mudanças nos padrões de uso e cobertura da terra em grande escala temporal e espacial. As imagens de satélites provenientes dos sensores remotos surgem como alternativa viável e eficiente, no estudo de informações sobre o uso da terra. Dessa forma, a análise de dados obtidos por sensores remotos de acordo com Jensen (2009) é de grande valor, pois tem como propósito extrair informações biofísicas da paisagem frequentemente.

Passo (2013) e Ruiz et al. (2014), destacam que as técnicas de Sensoriamento Remoto (SR) utilizam modernas plataformas aéreas e orbitais, nas quais são capazes de produzir imagens de alta resolução espacial, temporal e radiométrica. Essas técnicas permitem discriminar, de forma mais refinada, os diferentes alvos de interesse na superfície terrestre. Sendo assim, a disponibilidade de imagens com boa cobertura temporal e em diferentes resoluções espaciais, se tornou uma tarefa importante, no que diz respeito a identificação dos padrões de uso e cobertura da terra.

No entanto, a automação e o refinamento da análise dos padrões de uso e cobertura da terra podem ainda ser caracterizados como lacunas no monitoramento via SR, a qual é capaz de ser preenchida pelo uso de técnicas de aprendizado de máquina (AM). Nesse sentido, nos últimos anos métodos mais avançados como, por exemplo, os que utilizam algoritmos de aprendizado de máquina aliados às variáveis provenientes de séries temporais de sensores remotos, vem sendo cada vez mais utilizados.

O aprendizado de máquina pode auxiliar, de forma positiva, na análise dos padrões de uso e cobertura da terra, proporcionando modelos mais eficientes de processamento e interpretação. É importante salientar que o AM permite explorar de forma eficaz o potencial das imagens orbitais como estudo para diagnóstico ambiental, para a compreensão das implicações decorrentes de mudanças causadas pelas ações antrópicas e como apoio para decisões de ações ambientais, sociais, políticas e econômicas.

Cabe ainda destacar a importância de se utilizar o AM, na qual se dá pela construção do modelo ser de forma automatizada, onde os *softwares* podem encontrar informações em meio a grande quantidade de dados. Além disso, o

AM possui capacidade de aprender e responder de forma mais rápida e precisa, de identificar padrões e tomar decisões de acordo com a análise de dados baseadas em determinados algoritmos, com o mínimo da interferência humana.

O AM conforme descrito por Henriques (2008), permite a construção dos algoritmos baseados em métodos de classificação, onde esses algoritmos são capazes de automatizar o processo de atribuir determinadas informações da classe à qual pertence. Dessa forma, trabalhar com imagens de satélite vem se tornando cada vez mais importante e acessível, pois conforme Bunting et al. (2014), o avanço dos *softwares* e linguagens *open source*, bem como a implementação de algoritmos modernos, permitem estudar a cobertura da terra nas mais diversas escalas de tempo e espaço.

Pensando na evolução dos algoritmos de reconhecimento dos padrões do uso e cobertura da terra, os mais utilizados são a *Support Vector Machine* (CRISTIANINI e SHAWETAYLOR, 2001), *Random Forest*, *K-Nearest Neighbors*, *Naive Bayes* (CARVALHO et al., 2019). A utilização desses algoritmos se dá por ser de baixo tempo no processamento dos dados e, na maioria das vezes, apresentam elevada exatidão na identificação desses padrões.

Deste modo, visando conectar essas vertentes das tecnologias atuais, este trabalho teve como foco a utilização de algoritmos de aprendizado de máquina, a fim de analisar os dados obtidos de séries temporais dos sensores *Thematic Mapper* (TM) do Landsat 5 e *Operational Land Imager* (OLI) do Landsat 8, possibilitando averiguar o desempenho dos algoritmos e suas eficácias na análise dos padrões de uso e cobertura da terra.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

O objetivo da presente pesquisa constitui-se em analisar a capacidade dos algoritmos de aprendizado de máquina em prever o uso e cobertura da terra do município de Dona Francisca - RS.

1.1.2 Objetivos Específicos

1. Gerar imagens sintéticas dos sensores TM/Landsat 5 e OLI/Landsat 8 para tomada dos *pixels* de referência em diferentes datas (2004, 2009, 2015 e 2019);
2. Treinar os algoritmos de aprendizado de máquina *Random Forest* (RF), *K-Nearest Neighbors* (KNN), *Naive Bayes* (NB) e *Support Vector Machine* (SVM), a partir dos *pixels* de referência e;
3. Avaliar o desempenho dos algoritmos por meio do índice de Kappa e da acurácia global.

2 REVISÃO DE LITERATURA

2.1 SENSORIAMENTO REMOTO

O termo Sensoriamento Remoto (SR) se refere a utilização de modernos sensores, equipamentos para processamento e difusão de dados acomodados a bordo de aeronaves, espaçonaves, entre outras plataformas (NOVO, 2008). Segundo Moraes (2002) o SR permite obter informações dos objetos ou fenômenos que compõem a superfície terrestre sem estar em contato direto com o objeto sob investigação (Figura 1). Estes objetos ou fenômenos podem ser dividido em diferentes categorias, como vegetação, plantações, campo, água, solo exposto, estradas, entre outros aspectos, em uma área geográfica limitada na Terra.

Figura 1 - Processo de obtenção das imagens por meio do Sensoriamento Remoto.



Fonte: Florenzano (2011).

Loch (1993, p. 87) destaca que o SR se refere ao “conjunto de atividades, cujo objetivo reside na caracterização das propriedades de alvos naturais, através da detecção, registro e análise de fluxo de energia radiante, refletido ou emitido pelos mesmos”. No mesmo sentido Moraes (2002, p. 7) ressalta que estas atividades “envolvem a detecção, aquisição e análise (interpretação e

extração de informações) da energia eletromagnética emitida ou refletida pelos objetos terrestres e registradas por sensores remotos”, no qual esses sensores convertem a energia proveniente dos objetos em um registro em forma de imagem (ZANATA, 2014).

Esses sensores possuem diversos tipos de resolução, onde Florenzano (2007, p. 17) define resolução como a "capacidade de um sensor distinguir objetos da superfície terrestre". Dessa forma, as informações provenientes dos sistemas sensores remotos constituem características expressas pelas resoluções espacial, espectral, radiométrica e a resolução temporal (ROCHA, 2000).

Para Florenzano (2011) a resolução espacial, diz respeito a capacidade em diferenciar os objetos na superfície terrestre em função do seu tamanho sem perder a informação. Já para Crósta (1992), a resolução espacial está diretamente relacionada com o tamanho do *pixel* da imagem, isto é, uma determinada quantidade em metros quadrados de área no terreno que o sensor é capaz de imagear.

Benedetti (2010, p. 27) enfatiza a resolução espectral como “um conceito inerente às imagens multiespectrais, sendo definida pelo número de bandas espectrais de um sistema sensor e pela amplitude do intervalo de comprimento de onda de cada banda”.

A resolução radiométrica se refere a capacidade do sensor em distinguir a intensidade da energia refletida ou emitida pelos objetos, no que diz respeito aos valores associados ao nível de cinza utilizados para representar uma imagem (FORENZANO, 2011). Silva (1995), nessa mesma linha de entendimento, destaca que a resolução radiométrica representa a menor diferença de brilho que um sistema sensor é capaz de perceber, determinando se o alvo pode ser visto na imagem em função de seu contraste com alvos vizinhos.

Outras características dos sistemas sensores, é a resolução temporal no qual é dada em função do tempo que o satélite leva para voltar a recobrir a área de interesse, ou seja, é determinando o período mínimo a ser aguardado para um novo imageamento (BENEDETTI, 2010).

Assim, a variedade de sensores possibilita a aquisição de imagens com inúmeras características, as quais se adaptam às mais diversas aplicações do

Sensoriamento Remoto (PASSO, 2013), permitindo assim, inúmeras contribuições no desenvolvimento de metodologias a fim de fornecer dados sobre o uso e cobertura da terra no espaço geográfico considerado.

2.1.1 Satélites da série Landsat - *Land Remote Sensing Satellite*

O programa *Landsat Remote Sensing Satellite* (LANDSAT) consiste em um sistema de satélite de observação da Terra, que segundo Jensen (2009) é gerenciado pela Administração Nacional do Espaço e da Aeronáutica (*National Aeronautics and Space Administration* - NASA) e a Agência de Serviço Geológico dos Estados Unidos (*United States Geological Survey* - USGS).

O programa *Land Remote Sensing Satellite* de acordo com Moreira (2005) e Rocha (2000), teve o primeiro satélite da série foi o Satélite de Tecnologia dos Recursos Terrestres (*Earth Resources Technology Satellite – 1* (ERST – 1)) em 1972, e em janeiro de 1978 passou a ser chamado de Landsat 1. Este satélite conforme Cecconello (2017, p. 32), “possuía a bordo o sensor *Multispectral Scanner System* (MSS) e apresentava apenas quatro bandas, duas bandas na faixa do infravermelho e duas na faixa do visível apresentando uma resolução espacial de 80m e uma resolução temporal de 18 dias”. Já os satélites Landsat 2 e Landsat 3 lançados em 1975 e 1978 respectivamente, apresentavam as mesmas características do Landsat 1.

Em 1982 foi lançado o Landsat 4, porém agora passou a contar com o novo sensor *Thematic Mapper* (TM), sendo este projetado para dar suporte às pesquisas em diversas áreas destacando-se principalmente aos recursos naturais (SILVA, 2015; CECCONELLO, 2017).

O sensor *Thematic Mapper* (TM) do Landsat 5 foi difundido pela agência espacial norte-americana NASA e funciona em órbita equatorial a 705 km de altitude. O sensor imageador terrestre foi lançado em 1º de março de 1984 e permaneceu ativo até novembro de 2011, estabelecendo-se como o sensor de observação da Terra com maior período de operação de 28 anos (NASA, 2019).

Segundo a Embrapa (2013), o sensor TM contribui para diversos estudos que envolvem mapeamentos temáticos, principalmente na área de recursos naturais. A Tabela 1 demonstra as principais características do sensor TM/Landsat 5.

Tabela 1 - Características do sensor *Thematic Mapper* (TM) a bordo do satélite Landsat 5.

Sensor	Bandas Espectrais	Res. Espectral (μm)	Res. Esp.	Res. Temp.	Área Imag.	Res. Rad.
TM (<i>Thematic Mapper</i>)	(B1) Azul	0.45-0.52	30 m	16 dias	185 km	8 bits
	(B2) Verde	0.52-0.60				
	(B3) Vermelho	0.63-0.69				
	(B4) Infravermelho próximo	0.76-0,90				
	(B5) Infravermelho Médio	1.55-1.75				
	(B6) Infravermelho Termal	10.4-12.5	120 m			
	(B7) Infravermelho Médio	2.08-2.35	30 m			

Nota: Res.= Resolução, Esp.= Espacial, Temp.= Temporal, Imag.= Imageada, Rad.= Radiométrica.

Fonte: Embrapa (2013).

Os resultados promissores alcançados pela série Landsat foram tão expressivos que a NASA deu continuidade à série, e lançou o último satélite o Landsat 8, sendo lançado em 11 de fevereiro de 2013 e opera com os instrumentos Imageador Operacional da Terra (*Operational Land Imager - OLI*) e o Sensor de Infravermelho Termal (*Thermal Infrared Sensor - TIRS*) (ROY et al., 2014).

GANEM (2017, p. 49) enfatiza que os principais requisitos do Landsat 8 é ter,

Uma vida útil de cinco anos; uniformidade das larguras das bandas espectrais e dos comprimentos de onda centrais; desempenho radiométrico incluindo incerteza de calibração absoluta, relações sinal-ruído, sensibilidade à polarização e estabilidade; distâncias da amostra de solo e resposta da borda; geometria de imagem e geolocalização, incluindo o co-registo de bandas espectrais, sem falar na entrega de algoritmos de processamento de dados.

Na Tabela 2 são exibidas as bandas do Landsat 8/OLI com suas respectivas descrições, bem como, os dois tipos de sensores: Óptico, denominado *Operational Land Imager* (OLI) e o outro termal, denominado de *Thermal InfraRed Sensor* (TIRS).

Tabela 2 - Características do sensor *Operational Land Imager* (OLI) e *Thermal InfraRed Sensor* (TIRS) a bordo do satélite Landsat 8.

Sensor	Bandas Espectrais	Res. Espectral (μm)	Res. Esp.	Res. Temp.	Área Imag.	Res. Rad.
OLI (<i>Operational Land Imager</i>)	(B1) Costal	0.433-0.453	30 m	16 dias	185 km	16 bits
	(B2) Azul	0.450-0.515				
	(B3) Verde	0.525-0.600				
	(B4) Vermelho	0.630-0.680				
	(B5) Infravermelho Próximo	0.845-0.885				
	(B6) Infravermelho Médio	1.560-1.660				
	(B7) Infravermelho Médio	2.100-2.300				
	(B8) Pancromático	0.500-0.680	15 m			
	(B9) Cirrus	1.360-1.390	30 m			
TIRS (<i>Thermal Infrared Sensor</i>)	(B10) Infravermelho termal	10.60-11.19	100 m			
	(B11) Infravermelho termal	11.50-12.51	100 m			

Nota: Res.= Resolução, Esp.= Espacial, Temp.= Temporal, Imag.= Imageada, Rad.= Radiométrica.

Fonte: Embrapa (2013).

Segundo Oliveira (2017, p. 40), as imagens geradas no Landsat 8 possuem:

Um ciclo de repetição de 16 dias com resolução radiométrica de 16 bits. Os dados são coletados por meio de 11 bandas espectrais cuja resolução espacial é de 30m, com exceção da banda pancromática, que possui resolução de 15 m e a banda infravermelho termal que possui resolução de 100 m.

O Landsat 8, de acordo com Irons et al. (2012), tem como finalidade coletar, arquivar, processar e distribuir dados científicos de forma consistente com o registro de dados existente dos demais sensores da série Landsat.

2.1.2 Classificação digital e validação

A classificação digital de imagens é obtida por meio de *softwares* específicos, no qual consiste em associar a cada *pixel* um determinado objeto, visando identificar de forma eficiente os padrões e objetos de uso e cobertura da terra (ZHAI et al., 2012).

De acordo com Santos e Oliveira (2016), a classificação *pixel a pixel*, é baseada no comportamento espectral dos alvos sendo uma importante forma de obter informações sobre a superfície terrestre. Portanto, para classificar uma imagem em classes de interesse segundo Meneses e Sano (2012) o analista deve escolher pequenas áreas de amostras na imagem, contendo diversos *pixels* que descrevem um objeto real da superfície terrestre, de padrões, feições dos alvos por ele reconhecidos, ou ainda por meio de dados coletados à campo ou de mapas.

Quanto a análise da exatidão, essa é obtida por meio da medição da quantidade de acertos e erros do resultado da classificação ou da medição da acurácia proveniente do processo de atribuição dos *pixels* a determinadas classes, nas quais são provenientes do grau de confusão entre as classes em uma matriz de confusão ou também denominada de matriz de erros e pelo cálculo de índices, que fornecem uma medida geral, como o Índice Kappa (CONGALTON e GREEN, 1993; VIEIRA, 2000; KAWASHIMA, 2015).

KAWASHIMA (2015, p. 34) destaca que,

Uma matriz de erros ou de confusão é um arranjo de números estabelecidos em linhas e colunas que expressam o número de unidades de amostra (ou seja, *pixels*, aglomerados de *pixels* ou polígonos atribuído a uma determinada categoria em relação à

categoria real, está representada pela verdade verificada em terreno ou uma imagem de referência.

A Tabela 3 exemplifica a matriz de confusão, onde as colunas representam os dados de referência, enquanto as linhas indicam a classificação gerada a partir dos dados de Sensoriamento Remoto.

Tabela 3 - Matriz de confusão.

Classificação	Dados de Referência					Total linha
	Classes	1	2	3	k	
1	X _{1,1}	X _{1,2}	X _{1,3}	X _{1,k}	X ₁₊	
2	X _{2,1}	X _{2,2}	X _{2,3}	X _{2,k}	X ₂₊	
3	X _{3,1}	X _{3,2}	X _{3,3}	X _{3,k}	X ₃₊	
k	X _{k,1}	X _{k,2}	X _{k,3}	X _{k,k}	X _{k+}	
Total coluna	X ₊₁	X ₊₂	X ₊₃	X _{+k}	N	

Fonte: Adaptada de Iwai e Quintanilha (2005) e Congalton (1991).

O Índice Kappa é um coeficiente baseado na diferença entre a concordância observada (diagonal principal da matriz de confusão com a concordância entre a classificação e os dados de referência) e a chance de concordância entre os dados de referência de uma classificação aleatória (produto entre os totais marginais da matriz) (CONGALTON e BIGING, 1992).

Para Brites et al. (1996) o Índice Kappa (Equação 1) é indicado como medida de acurácia por representar toda a matriz de confusão, e não apenas a diagonal principal como é o caso da exatidão global (Equação 2).

$$K = \frac{N \sum_{i=1}^r X_{ij} - \sum_{i=1}^r X_{i+} X_{+i}}{N^2 - \sum_{i=1}^r X_{i+} X_{+i}} \quad (\text{Eq. 1})$$

Onde:

r = número de classes;

X_{ij}= número de elementos classificados corretamente; X_{i+}= total de elementos classificados para uma categoria *i*;

X_{+i}= total de elementos de referência amostrados para uma categoria *i*;

N= número total de amostras.

$$\text{Acurácia global: Classe K} = \frac{X_{1,1} + X_{2,2} + X_{3,3} + X_{k,k}}{N_{k,k}} \quad (\text{Eq. 2})$$

Os valores do Índice de Kappa variam entre 0 a 1, o qual os valores mais próximos de 1 indicam dados acurados. Esses valores são comparados e enquadrados segundo os conceitos definidos por Landis e Koch (1977), que servem para avaliar a exatidão da qualidade da classificação (Tabela 4).

Tabela 4 - Qualidade da classificação de acordo com os intervalos do coeficiente Kappa.

Valor do Kappa	Concordância
< 0,00	Péssima
0,0 – 0,20	Ruim
0,20 – 0,40	Razoável
0,40 – 0,60	Boa
0,60 – 0,80	Muito boa
0,80 – 1,00	Excelente

Fonte: Landis e Koch (1977).

2.2 GEOPROCESSAMENTO E SISTEMA DE INFORMAÇÃO GEOGRÁFICA

O Geoprocessamento conforme Inpe (2019) trata-se de um conjunto de tecnologias no qual está voltada para a coleta e tratamento de informações espaciais. Para Moreira (2005, p. 256), pode ser entendido como a “atualização de técnicas matemáticas e computacionais para tratar dos dados obtidos dos objetos ou fenômenos geograficamente identificados ou extrair informações desses objetos ou fenômenos, quando eles são observados por um sistema sensor”.

Já para Rosa (2013, p. 59) o Geoprocessamento é um “conjunto de tecnologias destinadas à coleta e tratamento de informações espaciais, assim como o desenvolvimento de novos sistemas e aplicações, com diferentes níveis de sofisticação”.

Ao abordar o tema Geoprocessamento, se faz necessário conceituar seus instrumentos computacionais, denominado de Sistema de Informação Geográfica (SIG), no qual compreende um conjunto de ferramentas

computacionais compostos por *software* e *hardware*. O SIG é usado para armazenar grandes quantidades de informação geográfica em banco de dados, além de analisar, manipular, interpretar, monitorar e disponibilizar informações georreferenciadas (LONGLEY, 2013; COSME, 2012).

Câmara et al. (2003, p. 94) destacam que o SIG “oferecem ferramentas que permitem a expressão de procedimentos lógicos e matemáticos sobre as variáveis georreferenciadas com uma economia de expressão e uma repetibilidade impossível de alcançar em análises tradicionais”. Calijuri e Cunha (2013) conceituam os SIG como ferramentas poderosas para análise e gerenciamento de informações de qualquer natureza espacial ou geográfica.

Hartmann et al. (2010, p. 11) consideram o SIG como sendo,

A principal ferramenta computacional para o Geoprocessamento, que nos permite definir as informações, as características e/ou atributos que variam em função de sua posição (espacial) e tempo (temporais). Esta ferramenta nos permite estudar a variabilidade dos atributos através da realização de análises complexas, ao integrar dados espaciais de diversas fontes de dados cartográficos (cartas, imagens, plantas) e dados de banco de dados georreferenciados, que através de fórmulas e/ou algoritmos, pesquisa os relacionamentos entre estas entidades gráficas e o banco de dados (informações pontuais), e da geoestatística.

Câmara e Queiroz (2004), destacam que existem três maneiras de utilizar um SIG, sendo para produção de mapas; suporte para análise espacial de fenômenos e; banco de dados geográficos. Numa visão mais abrangente pode-se dizer que um SIG é composto de cinco componentes independentes, no entanto, ligados uns aos outros por funções específicas, sendo: a interface com o usuário; entrada e integração de dados; funções de consultas e análise espacial; visualização e plotagem e; banco de dados geográficos (MOREIRA, 2011).

2.3 APRENDIZADO DE MÁQUINA

Os computadores em um determinado período eram apenas capazes de processar e efetuar o que lhes eram instruídos a fazer. Com o uso crescente das tecnologias, novas técnicas computacionais e ferramentas, foram surgindo com objetivo de possibilitar os computadores a aprender a se auto programar,

característica essa do campo do aprendizado de máquina (AM) proveniente do termo inglês, *Machine Learning* (ML).

Para Goldschmidt (2010), o aprendizado de máquina é uma área da inteligência artificial (IA), no qual tem como foco desenvolver técnicas computacionais que possuem capacidade de aprender por meio dos algoritmos, podendo assim, prever resultados, tendências e padrões.

O termo aprendizado de máquina refere-se à detecção automatizada de significativos padrões de dados, ou seja, são algoritmos que aprendem interativamente a partir de conjunto de dados fornecidos, gerando resultados preditivos ou descritivo com base no que aprendeu (SHWARTZ-SHALEV e BEN-DAVID, 2014). As técnicas de AM conforme Esteves et al. (2009) são empregadas na indução, com base em um conjunto de exemplos de um classificador, que deve ser capaz de prever a classe de novos dados quaisquer do domínio no qual ele foi treinado.

De acordo com Wang et al. (2009), o AM é o estudo da utilização de computadores para simular atividades humanas de aprendizagem e desenvolver métodos auto incrementais para obter novos conhecimentos, novas habilidades e identificação de conhecimentos já existentes. Ainda nesse sentido para Alpaydin (2010), o AM é o desenvolvimento de programas de computadores que são capazes de otimizar o seu desempenho utilizando dados de exemplos ou de experiência anterior.

Mitchell (1997) destaca que o AM visa treinar um conjunto de dados e testar estratégias que sejam mais eficazes para a construção de programas que possam aprender a partir da experiência, obtendo assim o conhecimento de forma automática. Técnicas de AM são cada vez mais empregadas para tratar problemas, por sua capacidade de aprender automaticamente por meio de grandes volumes de dados e produzir hipóteses úteis (MONARD e BARANAUSKA, 2003; CRACKNELL e READING, 2014).

Os algoritmos de AM segundo Carvalho (2011), aprendem a induzir uma determinada função ou hipótese, que seja capaz de resolver um problema baseado em dados que representam instâncias do problema a ser resolvido. Ou seja, é um método de análise de dados que desenvolve modelos, por meio do uso de algoritmos que aprendem interativamente a partir de entradas de dados,

possibilitando que os computadores encontrem resultados preditivos com base no que aprendeu.

O AM oferece uma vasta gama de opções de algoritmos diferentes, com a capacidade de classificar conjuntos de elementos (TAKAKURA et al., 2018). Como exemplo de AM pode-se citar os que representam analogias aos sistemas biológicos humano, como as Redes Neurais Artificiais (RNAs) e Algoritmos Genéticos, sistemas de aprendizado simbólico (Árvores de Decisão) e as teorias estatísticas como *Support Vector Machines* (SVMs) e *Random Forest* (MITCHELL, 1997; LARY et al., 2016).

Segundo Carvalho (2011), os algoritmos são utilizados em diversas tarefas e podem ser divididos em: descritivo, que consiste em explorar ou descrever um conjunto de dados e o preditivo, que visa encontrar uma função a partir dos dados de treinamento que possa ser usado para prever valores desconhecidos ou futuros de outras variáveis.

Os classificadores baseados em aprendizado de máquina, na classificação de imagens de satélite, vêm mostrando vantagens em relação aos classificadores supervisionados fundamentados em conceitos estatísticos, sendo que para a sua aplicação nenhum conhecimento da forma da distribuição de probabilidade dos dados de entrada é necessário face à maior flexibilidade que esses modelos possuem quanto à entrada de dados para a classificação (BISCHOF et al., 1992; RIBEIRO et al., 2005; GELETE e VOLATÃO, 2007).

Atualmente, diversos estudos têm demonstrado a importância do uso de algoritmos de aprendizado de máquina para classificação de alvos a partir de imagens de SR (CARVALHO et al., 2019) devido à alta performance preditiva em análise de dados. Os algoritmos supervisionados têm sido usados com intuito de melhorar a classificação do uso e cobertura da terra, como por exemplo, a *Support Vector Machines* (SVM), *Random Forest* (RF), *Naive Bayes* (NB) e o *K-Nearest Neighbor* (KNN). Conforme Li et al. (2014), os algoritmos têm demonstrado excelente desempenho na análise de bases de dados do Sensoriamento Remoto. A SVM, por exemplo, é empregada para compreender problemas de classificação de dados provenientes do SR (MERCIER et al., 2003; GONÇALVES et al., 2006) e os “resultados mostram que estas podem obter uma melhor performance na classificação e ter melhor capacidade de generalização do que classificadores estatísticos tradicionais” (SOUSA, 2009, p. 15). Zhu et al.

(2012), sugerem o algoritmo SVM para avaliar as variações ocorridas na cobertura da terra de uma determinada área.

Referente ao algoritmo RF, esse é um classificador estatístico muito eficaz e com vasta aplicação, sendo este eficiente na classificação de imagens de satélite (CUTLER et al., 2007; GISLASON et al., 2006). Oliveira (2018, p. 15) ressalta que o RF vem sendo cada vez mais “utilizado na classificação do uso do solo de imagens de sensores remotos por produzir classificações excelentes e apresentar tempos de processamento interessantes na classificação de imagens”. Os algoritmos NB e KNN também são comumente usados e têm sido aplicados em diversos métodos na classificação de dados provenientes do SR, como pode ser visto nos trabalhos desenvolvidos por Moura (2018), Santos (2014), Li et al. (2014) e Albuquerque (2012).

2.3.1 Modelos supervisionados e não supervisionados

A aprendizagem de máquina pode ser classificado em dois tipos: supervisionados quando o objetivo é a resolução de problemas de classificação ou regressão, enquanto que, o não supervisionado é dividido em agrupado ou associado (MONARD e BARANAUSKAS, 2003). De forma sucinta, Silva et al. (2017, p. 69-70) enfatizam que a “diferença básica entre esses dois modos de aprendizagem, é a presença ou não do atributo que rotula os exemplos do conjunto de dados”.

O aprendizado supervisionado engloba a construção de um modelo estatístico, no qual prevê ou estima um resultado baseado em uma ou mais entradas de dados classificados (MONARD e BARANAUSKAS, 2003). Silva et al. (2017, p. 70), ressaltam que o aprendizado supervisionado tem como objetivo “induzir um classificador (ou hipótese), por meio de um conjunto expressivo de dados previamente rotulados, para classificar novos exemplos ainda não rotulados”.

Esta forma de aprendizagem supervisionada para Montañó, (2016) é utilizada em situações onde se conhece os atributos de saída dos dados de treinamento, o que permite avaliar a capacidade do resultado em prever prováveis situações futuras.

Referente ao aprendizado não supervisionado a função a ser aprendida não está explícita, ou seja, não se conhece os atributos de saída, e os algoritmos devem ter a capacidade de organizar e aprender funções os quais se referem os dados, baseados em agrupamentos e vizinhança (SIVIERO e JÚNIOR, 2019).

2.3.2 K-Vizinhos mais próximos (*K-Nearest Neighbors* - KNN)

O algoritmo *K-Nearest Neighbor* (KNN) é um método supervisionado de classificação de dados, no qual é empregado conforme destaca Xu et al. (2013) no reconhecimento de padrões, baseando-se na técnica do vizinho mais próximo (*Nearest Neighbor*), onde 'k' é o número de vizinhos a serem considerados no momento da classificação. Uma forma de definir o vizinho mais próximo a cada ponto no espaço é por meio do cálculo da distância Euclidiana (MITCHELL, 1997) sendo representado pela Equação 3:

$$d(x_i, x_j) \equiv \sqrt{\sum_{r=1}^n (a_r(x_i) - a_r(x_j))^2} \quad (\text{Eq. 3})$$

Onde:

$d(\vec{x}_i, \vec{x}_j)$ = distância entre os pontos \vec{x}_i e \vec{x}_j ;

$a_r(\vec{x}_i)$ = é a característica r da instância \vec{x}_i ;

n = é o número total de coordenadas de cada ponto.

O objetivo do KNN é formar uma generalização com base em um conjunto de treinamento, maximizando a acurácia da classificação de novos dados (GAO e GAO, 2010). O algoritmo de acordo com Su (2011) presume que o conjunto de treinamento seja composto pelas variáveis descritivas e pela sua classificação, então o KNN utiliza essas variáveis para classificar um novo item.

Dados das amostras dos *pixels* das imagens de Sensoriamento Remoto estão distribuídas no espaço de características e o algoritmo KNN usa uma medida de distância para definir a proximidade de uma amostra com a outra. Esse algoritmo classifica novas amostras conforme as "K" amostras mais próximas do conjunto de treinamento (DUDA et al., 2000; SANTOS, 2013;

CRACKNELL e READING, 2014; SARMIENTO et al., 2014). Segundo Lin et al. (2012), o KNN têm sido utilizado com sucesso para a classificação de dados de séries temporais.

2.3.3 Máquinas de Vetores de Suporte (*Support Vector Machine – SVM*)

O *Support Vector Machine* (SVM) de acordo com Vidhya (2013) é um grupo de métodos de aprendizado de máquina usado para classificação e regressão, no qual está baseado no conceito de planos de decisão que definem os limites de decisão, estes planos separam objetos de diferentes classes. Para Cortes e Vapnik (1995); Adam et al. (2014), o SVM é um método de aprendizado de máquina supervisionado treinado para encontrar o hiperplano ótimo, OSH (do inglês *Optimal Separating Hyperplane*) que é capaz de diferenciar seus elementos por meio de diferentes amostras (classes).

O processo de treinamento do algoritmo SVM segundo Nascimento et al. (2009, p. 2080) consiste em:

“treinar um classificador de forma que este aprenda um mapeamento $x \mapsto y$ por meio de exemplos (classes) de treinamento $\{x_i, y_i\}$ de forma que a máquina seja capaz de classificar um exemplo (x, y) ainda não visto que siga a mesma distribuição de probabilidade (P) dos exemplos de treinamento”.

Para Smola et al. (2000), o algoritmo SVM possui vantagens como: ter uma teoria bem estabelecida dentro da matemática e estatística; apresenta robustez em dados de grandes dimensões como, por exemplo, imagens de satélite; ótima capacidade de generalização, ou seja, apresenta eficiência para prever dados que não pertençam ao conjunto utilizado no treinamento.

O SVM por apresentar um bom desempenho tornou-se um referencial na pesquisa de aprendizagem de máquina (SOLIMAN e MAHMOUD, 2012). Nos últimos anos a técnica SVM vem sendo utilizada em inúmeras tarefas de reconhecimento de padrões, obtendo resultados que demonstram características promissoras como a boa capacidade de generalização intrínseca e a robustez ao ruído (ALI et al., 2015).

2.3.4 Florestas aleatórias (*Random Forest* – RF)

O classificador *Random Forest* (RF) (Floresta Aleatória), foi desenvolvido por Breiman (2001) no qual designa-se como sendo uma técnica de AM que pode ser usada para regressão e classificação. Ainda de acordo com o autor, o RF é um algoritmo baseado na combinação de diversos modelos de forma a aumentar a acurácia da classificação.

O termo *Random Forest* se originou pelo fato de:

Se tratar de um conjunto de árvores de decisões, e seu funcionamento se explica com a separação do grupo original de dados em diversos subconjuntos (cada grupo de subconjunto pode ser considerado como uma árvore de decisão) e a partir daí novas árvores são geradas a partir da ruptura de cada subconjunto (JÚNIOR, 2018, p. 36).

Para Neto (2014), o *Random Forest* tem como finalidade efetuar a criação de várias árvores de decisão, de tal forma que cada árvore é usada um subconjunto dos atributos selecionados aleatoriamente a partir do conjunto de dados originais, ou seja, o RF divide a amostra em inúmeras amostras menores, e a partir de cada uma dessas pequenas amostras seleciona exemplos de uma maneira específica, chamado de *bootstrapping*, possibilitando assim melhor análise dos dados.

O algoritmo RF é um classificador que faz uso do método de árvore de decisão, no entanto, se diferencia um pouco dos algoritmos de Árvores de Decisão, pois consiste em:

Técnica de agregação de classificadores do tipo árvore de decisão, construídos de forma que sua estrutura seja composta de maneira aleatória. Para determinar a classe de uma instância, o método combina o resultado de várias árvores de decisão, por meio de um mecanismo de votação (LORENZETT e TELÖCKEN, 2016, p.1).

Conforme Belgiu e Drăgu (2016), a RF é o classificador amplamente utilizado com dados de Sensoriamento Remoto para a classificação do uso e ocupação do solo, seja utilizando dados multiespectrais ou hiperespectrais.

2.3.5 Naive Bayes – NB

O *Naive Bayes* (NB) é um conjunto de algoritmos de aprendizado supervisionado baseado na aplicação do Teorema de *Bayes*, estando fundamentado no modelo de classificação probabilística de aprendizado de máquina, no qual assume que cada atributo influenciará de forma independente para cada nova instância a ser classificada (WIELAND e PITTORE, 2014; AMARAL, 2016).

O teorema de Bayes segundo Domingos e Pazzani (1996) realiza e substancia o desenvolvimento de um procedimento, no que tange o aprendizado de máquina, podendo ser usado para a predição da classe associada à uma nova instância, dado um conjunto de instâncias de treinamento. A Equação 4, para Cavalcanti (2016) permite calcular a probabilidade de um evento (A) dado que outro evento aconteceu (B).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (\text{Eq. 4})$$

Sendo que: a probabilidade isolada de que um evento A (probabilidade a priori) aconteça é expressa por $P(A)$.

A probabilidade a posteriori de um evento A dado que ocorreu também um evento B é expressa por $P(A|B)$ (CAVALCANTI, 2016).

O NB é um algoritmo eficiente em diferentes aplicações e nos últimos anos tem sido usado por pesquisadores em classificações. De acordo com Chen e Sycara (1997), o NB apresenta facilidade na construção do seu algoritmo, o processo de classificação é muito eficiente no que diz respeito quando os atributos são independentes entre si, além de ser rápido na aprendizagem e predição (FRIEDMAN et al., 1997).

2.4 SOFTWARE LIVRE PARA SENSORIAMENTO REMOTO

2.4.1 *Konstanz Information Miner* – KNIME

O *software* KNIME, abreviação de *Konstanz Information Miner* é uma ferramenta de mineração desenvolvida em 2004 pela Universidade de Konstanz, em Konstanz, Alemanha, e em 2006 foi lançada a primeira versão desse minerador. O KNIME é uma plataforma de análise preditiva de *software* livre, *open source*, tendo a licença *General Public License* (GLP) (KNIME, 2019).

De acordo com Borges (2016), o KNIME permite o desenvolvimento de estatísticas sofisticadas e operações de *data mining* e *machine learning*, de forma a realizar o pré-processamento, classificação, regressão, segmentação, associação, visualização, análise preditiva e a geração de relatórios.

O fluxo de criação e execução de *workflow* no KNIME é baseado na ideia de adição de “nodes” (nó), onde estão vinculados por meio de suas entradas e saídas. Segundo Feltrin (2015), cada nó executa processamento baseado em algoritmo de dados, no qual é capaz de interagir com outros nós (quando conectados entre si), permitindo assim, gerar e gravar os fluxos de trabalho de processamento de dados complexos.

Cabe destacar que em cada nó existe um semáforo abaixo dele, onde indica o *status* de prontidão desse nó. Se estiver na tonalidade verde, isso significa que o nó já foi executado e concluiu seu trabalho, se estiver amarelo, o nó está devidamente configurado e pronto para ser executado, porém, ainda não concluiu seu trabalho, por fim se estiver vermelho, o nó ainda não está configurado e não pode ser executado até que tenha sido configurado corretamente. Dessa forma, o KNIME, por meio da sua abordagem intuitiva, facilita a aplicação de ferramentas complexas de aprendizado de máquina.

O KNIME foi projetado nos seguintes princípios:

Estrutura visual e interativa, onde os fluxos de dados devem ser combinados arrastando e soltando uma variedade de unidades de processamento; modularidade, as unidades de processamento e contentores de dados não devem depender um do outro a fim de facilitar a distribuição e permitir o desenvolvimento independente de diferentes algoritmos; e fácil expansão, deve ser fácil adicionar novos nós de processamento ou visualizações e distribuir por meio de um simples mecanismo de *plugins* sem a necessidade de procedimentos

complicados de instalação e desinstalação (BERTHOLD et al., 2008, p. 36).

O *software* KNIME vem sendo utilizado em diversas áreas, como por exemplo, em estudos sobre a qualidade da cimentação de poços de petróleo (isolamento hidráulico por meio de perfis acústicos) (LIMA, 2018), classificação do solo (DHIMAN, 2017), previsão de duração e procedimentos cirúrgicos (CARVALHO, 2017), análise de dados de rede social (*Twitter*) e análise de sentimentos para o apoio à decisão (ANDRADE, 2015), na área da química (BRAGA, 2015; CAVALCANTI, 2018). Feltrin (2015), ainda destaca que 50% dos usuários são provenientes de campos bastante diferentes como o das ciências, medicina, negócios e ciências sociais.

Partindo dessa premissa, a escolha do *software* KNIME a ser usado nesta dissertação, se deu em razão dos seguintes requisitos:

1. ser um *software* livre;
2. de fácil entendimento e modelagem;
3. possui amplo material de apoio aos usuários;
4. capacidade de integrar diferentes linguagens de programação no mesmo ambiente de fluxo de trabalho e;
5. principalmente por quase não existir trabalhos voltados à área florestal, usando dados provenientes de séries temporais de sensores remotos.

2.4.2 Software QGIS

O *software* QGIS é um SIG, de código aberto e está disponível gratuitamente para uso. Atualmente o QGIS é uma aplicação SIG de fácil utilização que funciona em diferentes sistemas operacionais como o Linux, Unix, Mac OSX, Windows e Android, onde as suas funcionalidades podem ser estendidas por meio de inúmeros *plugins*, denominados também de complementos (NADER e FAGGIN, 2011).

O QGIS é um *software* robusto e versátil, caracterizado pela acessibilidade ao código fonte. Suporta diversos formatos geoespaciais como vetores, *rasters* e bases de dados (*SpatialLight* do *Open Source* e suporta

PostgreSQL/PostGIS) (FRIEDRICH, 2014), associado a interface minimalista e eficiente e com capacidade de exibição de camadas de servidores remotos (LOBO et al., 2012).

Steiniger e Hay (2009) destacaram que o QGIS é visto como uma das mais promissoras plataformas livres para SIG, tendo em vista a interface eficaz e intuitiva. Partindo dessa premissa, é notável uma demanda crescente em ambientes acadêmicos e profissionais, a utilização do QGIS, com propósito em reproduzir o conjunto exato dos dados e dos códigos utilizados em estudos voltados às ciências florestais.

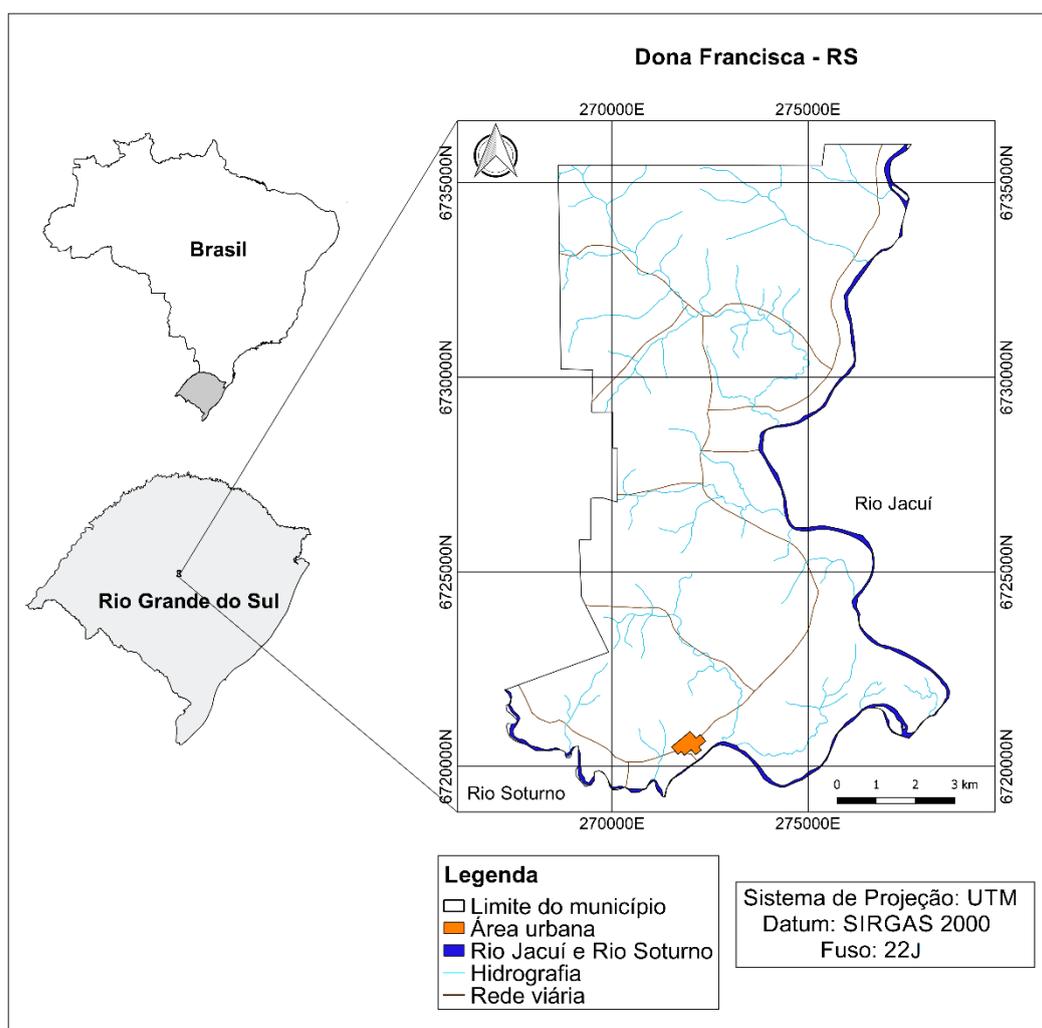
3 MATERIAL E MÉTODOS

3.1 CARACTERIZAÇÃO DA ÁREA DE ESTUDO

3.1.1 Localização da área de estudo

O município de Dona Francisca, pertence a região da Quarta Colônia e está localizado na Depressão Central do Estado do Rio Grande do Sul, estando distante a 210 Km da capital Porto Alegre. Possui como coordenadas de referência 29° 36' 41" de latitude Sul e 53° 21' 03" de longitude Oeste (Figura 2). A área territorial do município de acordo com o Ibge (2010) é de 114,346 Km².

Figura 2 - Localização do município de Dona Francisca, Rio Grande do Sul.



Fonte: Autora (2020).

3.1.2 Aspectos físicos e socioeconômicos

De acordo com a classificação de Köppen, o município de Dona Francisca possui o clima do tipo subtropical úmido – Cfa, onde os meses mais quentes vão de dezembro a fevereiro com temperatura média de 22°C, enquanto que os meses junho, julho e agosto são os mais frios, tendo uma temperatura média de -3°C e 18°C. A precipitação média anual é de aproximadamente 1.722 mm (ALVARES et al., 2013).

O relevo do município caracteriza-se por Planície alúvio-coluvionar, Depressão do Rio Jacuí, Serra Geral e o Planalto dos Campos Gerais (IBGE, 2019). A área acidentada faz parte da Serra Geral formada por sucessivos derrames basálticos, estando presentes o Morro Santo Antônio e o Morro Formoso (LAGO e FARENZENA, 2008). Geologicamente o município apresenta as seguintes formações: Depósito-aluvionares, Caturrita, Botucatu e Gramado (CPRM, 2019). A vegetação é composta por Floresta Estacional Decídua da borda do sul do Planalto Meridional Brasileiro, banhados, várzea e matas de galeria da Depressão Central (PMSBDF, 2011).

Os solos presentes no município segundo Embrapa (2013) são do tipo Planossolos Háplicos Eutróficos (Sxe13) e Neossolos Litólicos Eutróficos (RLe23). Streck et al. (2008), ressaltam que os Planossolos são solos imperfeitamente ou mal drenados, ocorrendo em áreas de várzea com relevo plano a suave ondulado. Os Neossolos são solos rasos, de formação recente e encontrados nas mais diversas condições de relevo e drenagem.

O município de Dona Francisca pertence à região hidrográfica do Guaíba e a bacia hidrográfica do Alto Jacuí, tendo como principais cursos d'água os rios Jacuí e Soturno, os quais são utilizados para o abastecimento de água, fornecimento de energia, pesca e irrigação. Conforme Favera (2006), o sistema fundiário é baseado em pequenas propriedades rurais, e tem como a sua principal base econômica a agricultura, com destaque para a produção do arroz irrigado nas áreas planas e ao longo das margens dos rios Jacuí e Soturno. Quanto as culturas do fumo, feijão e milho são cultivadas em áreas de encostas de morro (VENDRUSCOLO, 2010).

Dados do Instituto Brasileiro de Geografia Estatística (2010), o município de Dona Francisca possui uma população residente de 3.401

habitantes, onde 2.146 são residentes da zona urbana e 1.255 da zona rural.

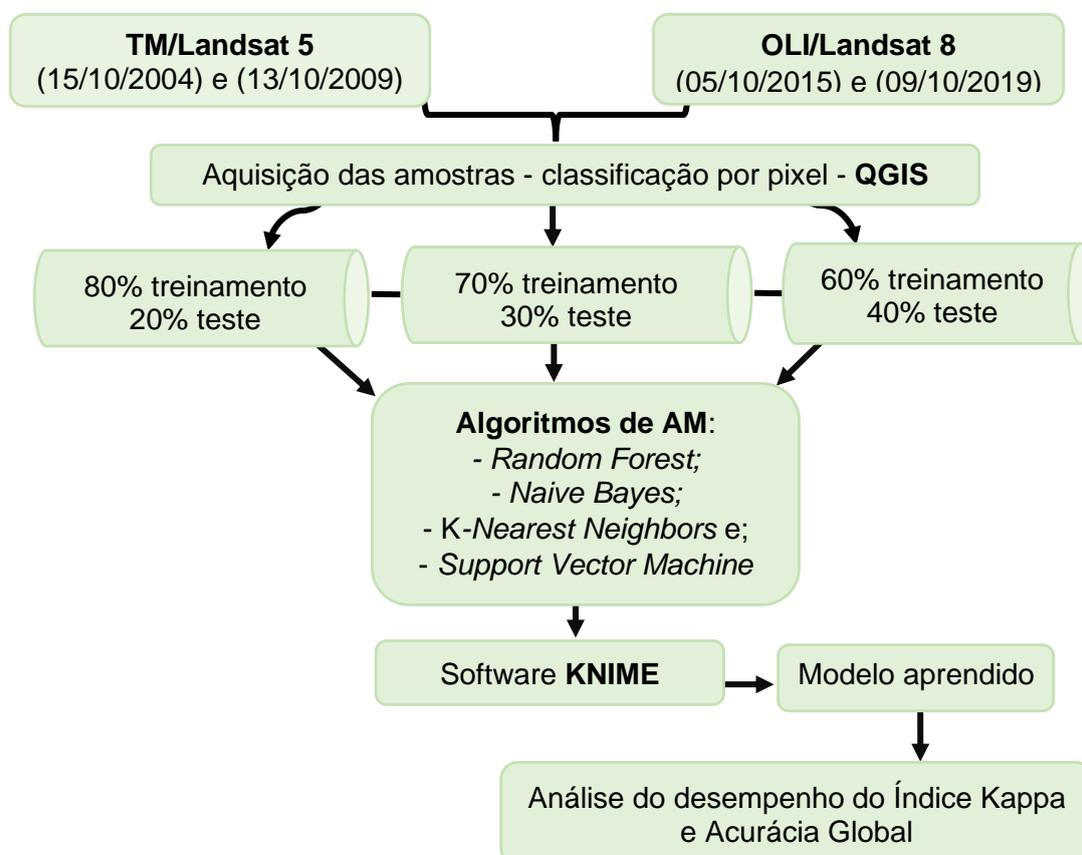
3.2 MATERIAL

No desenvolvimento da presente pesquisa foram utilizadas imagens orbitais dos sensores TM/Landsat 5 e OLI/Landsat 8; *shapefile* dos limites políticos municipais, na escala 1:50.000 (IBGE, 2019) e para o processamento dos dados utilizou-se os *softwares* QGIS versão 2.14 “Essen” e KNIME (Konstanz Information Miner) versão 4.0.

3.3 MÉTODOS

Os procedimentos metodológicos na Figura 3 representam as etapas cumpridas para alcançar os resultados da presente pesquisa.

Figura 3 - Fluxograma dos procedimentos metodológicos.



Fonte: Autora (2020).

3.3.1 Processamento dos dados no QGIS

O banco de dados geográfico foi construído no *software* QGIS versão 2.14 “Essen”, um *software* livre licenciado sob a *General Public License* (GLP). Para o mesmo foram importados os dados vetoriais do Estado do Rio Grande do Sul com os seus respectivos limites físicos municipais, a partir dos quais foi delimitada a área de interesse. A projeção cartográfica foi em UTM (Universal Transversa de Mercator) e Datum SIRGAS 2000, fuso 22 J. Posteriormente foram importadas as imagens do sensor TM/Landsat 5 e OLI/Landsat 8.

3.3.2 Processamento das imagens

As imagens dos sensores TM/Landsat 5 e OLI/Landsat 8 (Tabela 5) foram adquiridas gratuitamente na página do Instituto Nacional de Pesquisa Espacial (INPE) e no site do Centro de Sensoriamento Remoto do Serviço Geológico dos Estados Unidos (USGS), respectivamente.

Tabela 5 - Imagens abrangentes do TM/Landsat 5 e OLI/Landsat 8 e as respectivas datas, Dona Francisca – RS.

Cena	Órbita	Ponto	Data de aquisição	Satélite	Sensor
1	222	081	15/10/2004	Landsat 5	TM
			13/10/2009		
			05/10/2015	Landsat 8	OLI
			09/10/2019		

Fonte: Autora (2020).

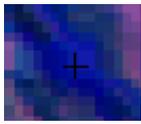
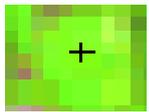
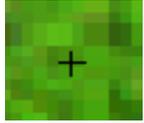
Para a aquisição das imagens levou-se em consideração a escolha de 5% de nuvens nas imagens, bem como a época, sendo selecionadas imagens do mesmo período, mês de outubro. A época de aquisição da imagem de acordo com Moreira (2005) é um fator importante e que contribui para o êxito da análise de dados de satélites.

Antes de iniciar o processamento das imagens foi necessário instalar o complemento *Semi-Automatic Classification Plugin* (SCP), na versão 4.9.2,

desenvolvida por Congedo (2019) e integrado ao software QGIS. Posteriormente, iniciou-se a importação das imagens TM/Landsat 5 (Bandas 1, 2, 3, 4, 5, 7) e OLI/Landsat 8 (Bandas 2, 3, 4, 5, 6, 7). Em seguida, criou-se o *Band set* que consiste em uma única imagem com todas as seis bandas, possibilitando obter as diferentes combinações que resultaram em distintos padrões de cores representativas do uso e cobertura da terra.

Partindo dessa premissa, realizou-se a amostragem dos *pixels* das imagens sintéticas dos sensores TM/Landsat 5 e OLI/Landsat 8 levando-se em consideração apenas a informação espectral de cada *pixel* (valor numérico), com a finalidade de obter informações para caracterizar e diferenciar os padrões de uso e cobertura da terra (Quadro 1).

Quadro 1 – Padrões de uso e cobertura da terra visualizada nas imagens RGB 543 do sensor TM/Landsat 5 e RGB 654 do sensor OLI/Landsat 8.

Classes	Características	Amostra Real
Água	Reservatórios naturais e/ou artificiais e, em alguns casos, áreas utilizadas para o cultivo do arroz.	
Agricultura	Áreas agrícolas cobertas por vegetação.	
Campo	Vegetação rasteira composta de gramíneas e pastagens.	
Floresta	Formações vegetais de porte arbóreo, dentre as quais florestas nativas e plantadas.	
Solo exposto	Áreas agrícolas em preparo ou pousio, voçorocas ou processos erosivos.	

Fonte: Autora (2020).

Em seguida, os dados gerados foram salvos em uma planilha de Excel®

na extensão *.csv. Nesta planilha foram salvas as coordenadas geográficas (X e Y), padrões de uso e cobertura da terra (água, campo, floresta e solo exposto) e as seis bandas espectrais (TM/Landsat 5 - Bandas 1, 2, 3, 4, 5 e 7 e OLI/Landsat 8 - Bandas 2, 3, 4, 5, 6 e 7), com seus respectivos valores numéricos (informação espectral de cada *pixel*).

3.3.3 Aprendizado de máquina no *software* KNIME

Por meio do conjunto das amostras provenientes dos *pixels* de referência das imagens de satélite para as diferentes datas foi possível preparar os dados para leitura, permitindo assim, iniciar o processo de aprendizado de máquina no *software* KNIME.

Para analisar o desempenho dos algoritmos, os modelos preditivos foram desenvolvidos com vistas para os padrões de uso e cobertura da terra, sendo adaptados com base nos trabalhos desenvolvidos por Dhiman et al. (2017); Antunes (2018); Basha et al. (2018) e na própria página do KNIME (2019). Os nós utilizados e o funcionamento dos mesmos no *software* KNIME estão descritos a seguir:

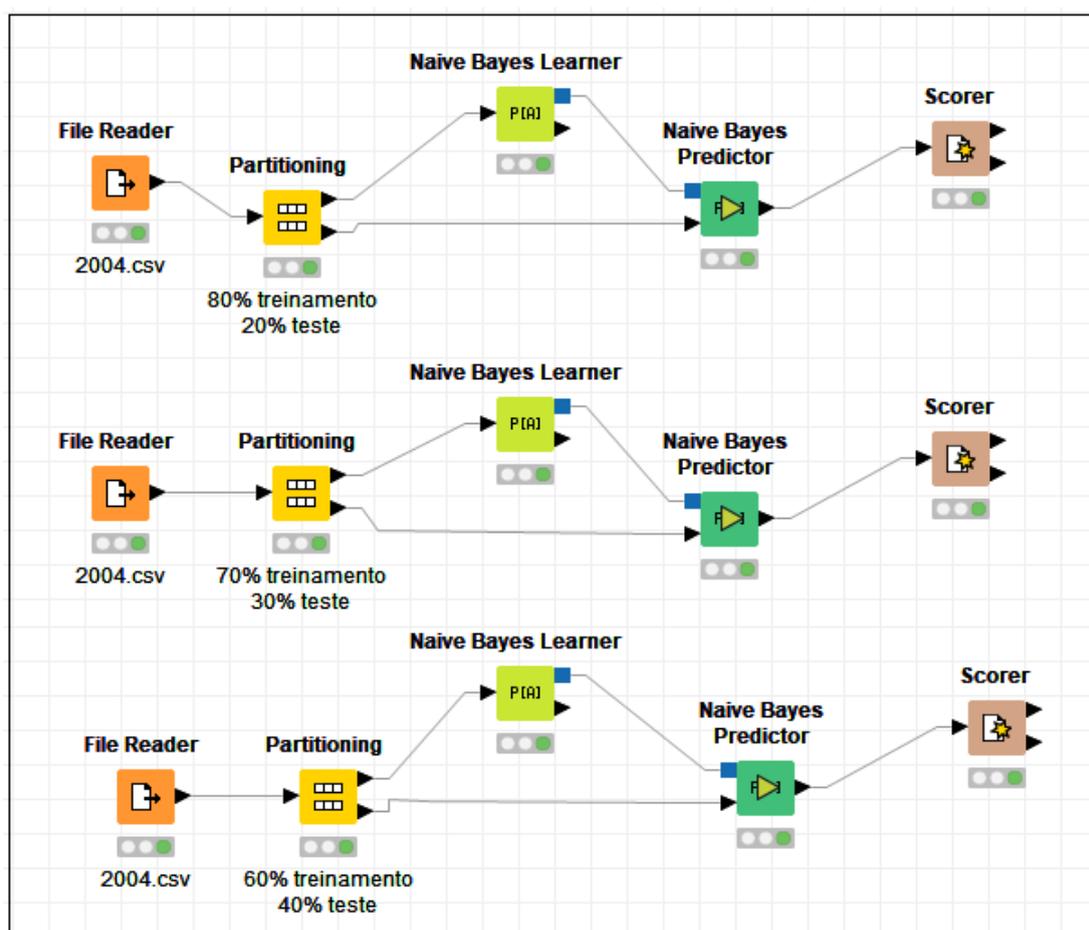
1. *File Reader*: responsável pela leitura dos dados de um repositório, neste caso a entrada do arquivo foi no formato *.csv;
2. *Partitioning*: a tabela de entrada é dividida em duas partições, onde esse nó é usado para produzir um conjunto de dados de treinamento (onde o modelo é gerado) e o conjunto de dados de teste (avaliação do modelo treinado);
3. *Naive Bayes*, *Random Forest*, *Support Vector Machine* e *K-Nearest Neighbors Learner*: estes nós criam um modelo a partir dos dados fornecidos;
4. *Naive Bayes*, *Random Forest*, *Support Vector Machine* e *K-Nearest Neighbors Predictor*: prevê a classe por linha com base no modelo aprendido e;
5. *Scorer*: mostra a matriz de confusão, acurácia global e o índice de Kappa de Cohen, entre outros dados.

As amostras dos *pixels* de referência para o treinamento e teste foram definidas e adaptadas com base nos autores Marsland (2008); Buduma e

Locascio (2015); Mueller e Massaron (2019) em 80% a 20%, 70% a 30% e 60% a 40%, respectivamente.

A construção do modelo preditivo, como por exemplo do algoritmo *Naive Bayes* (Figura 4), utilizando o *software* KNIME, mostra que o nó *File reader* é o leitor do arquivo que contém os dados dos *pixels* de referências dos padrões do uso e cobertura da terra.

Figura 4 - Exemplo do fluxo de processos e operadores usados no modelo preditivo do algoritmo *Naive Bayes*.



Fonte: Autora (2020).

O conjunto das amostras dos *pixels* de referência de cada data foram estratificadas usando o nó "Patittioning". O nó "Patittioning" divide as amostras em dados de treinamento e de teste, no qual permite averiguar se o modelo apresenta boa performance não apenas em dados utilizados para seu ajuste, mas também quanto à capacidade de generalização para novas observações. O nó *Partitioning* divide o arquivo em dois conjuntos de dados (80% para

treinamento e 20% para teste), essa divisão do arquivo foi realizada também para os demais treinamentos e teste de 70% e 30% bem como, para o treinamento e teste de 60% e 40%, respectivamente.

O nó *Naive Bayes Learner*, foi utilizado para criar um modelo bayesiano a partir dos dados de treinamento no qual é fornecido. Além disso, calcula o número de linhas por valor de atributo por classe para atributos nominais e a distribuição Gaussiana para atributos numéricos (KNIME, 2019). Referente ao nó *Naive Bayes Predictor*, foi usado para teste no modelo aprendido, modelo este criado anteriormente pelo nó denominado *Naive Bayes Learner*.

O último nó a ser adicionado no fluxo de trabalho é o *Scorer*, com intuito de averiguar o resultado do modelo criado *Naive Bayes Predictor*, aplicado ao conjunto de dados. O *Scorer* reporta uma série de estatística de precisão, como por exemplo, a acurácia global e o índice de Kappa de Cohen.

A predição conforme ressalta Silva (2006) é utilizada para definir um provável valor para as variáveis, os classificadores preditivos geram a probabilidade de uma instância pertencer a uma dada classe, para os quais diferentes limiares podem ser aplicados para gerar diferentes conjuntos de saída.

O fluxo de criação e execução do *workflow* demonstrado na Figura 4, foi realizado para os anos de 2004, 2009, 2015 e 2019, totalizando 12 processos para cada um dos algoritmos (*Naive Bayes*, *Random Forest*, *K-Nearest Neighbor* e o *Support Vector Machine*).

4 RESULTADOS E DISCUSSÃO

4.1 PIXELS DE REFERÊNCIA

Com base nos mapas resultantes da composição sintética das bandas RGB 543 do sensor TM/Landsat 5 e RGB 654 do sensor OLI/Landsat 8, foi possível aumentar a discriminação entre os elementos da superfície terrestre, identificando os padrões de uso e cobertura da terra, sendo elas: água, agricultura, campo, floresta e solo exposto. Após a composição e a identificação obteve-se as amostras dos *pixels* de referência dos padrões de uso e cobertura da terra.

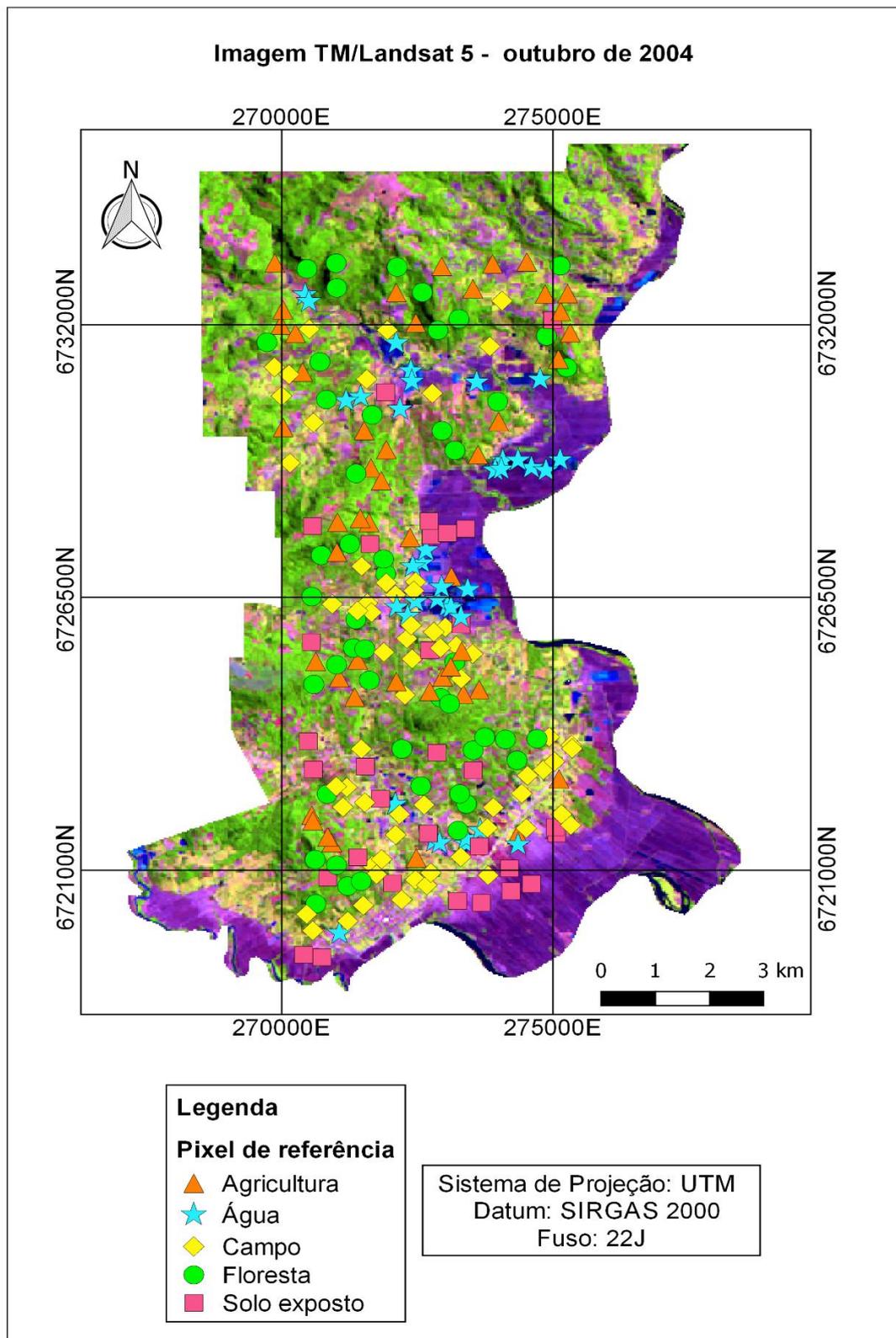
Dessa forma, com o resultado da composição sintética das imagens, as amostras dos *pixels* de referência dos padrões de uso e cobertura da terra do município de Dona Francisca foram adquiridas de forma aleatória, totalizando 250 *pixels* para cada série temporal nos anos de 2004, 2009, 2015 e 2019 conforme mostram as Figuras 5, 6, 7 e 8.

Faria (2015, p. 34-35), enfatiza que tamanho do “conjunto de amostras pode gerar impacto considerável na exatidão dos resultados obtidos, no entanto, não há um consenso acerca de qual tamanho deve ter o arquivo de treinamento em estudos que empregam classificação supervisionada”.

Segundo Fassnacht et al. (2014) e Ghosh et al. (2014), um pequeno número de amostras de 60 *pixels* por classe pode ser suficiente para o treinamento ideal da *Random Forest* e *Support Vector Machine*.

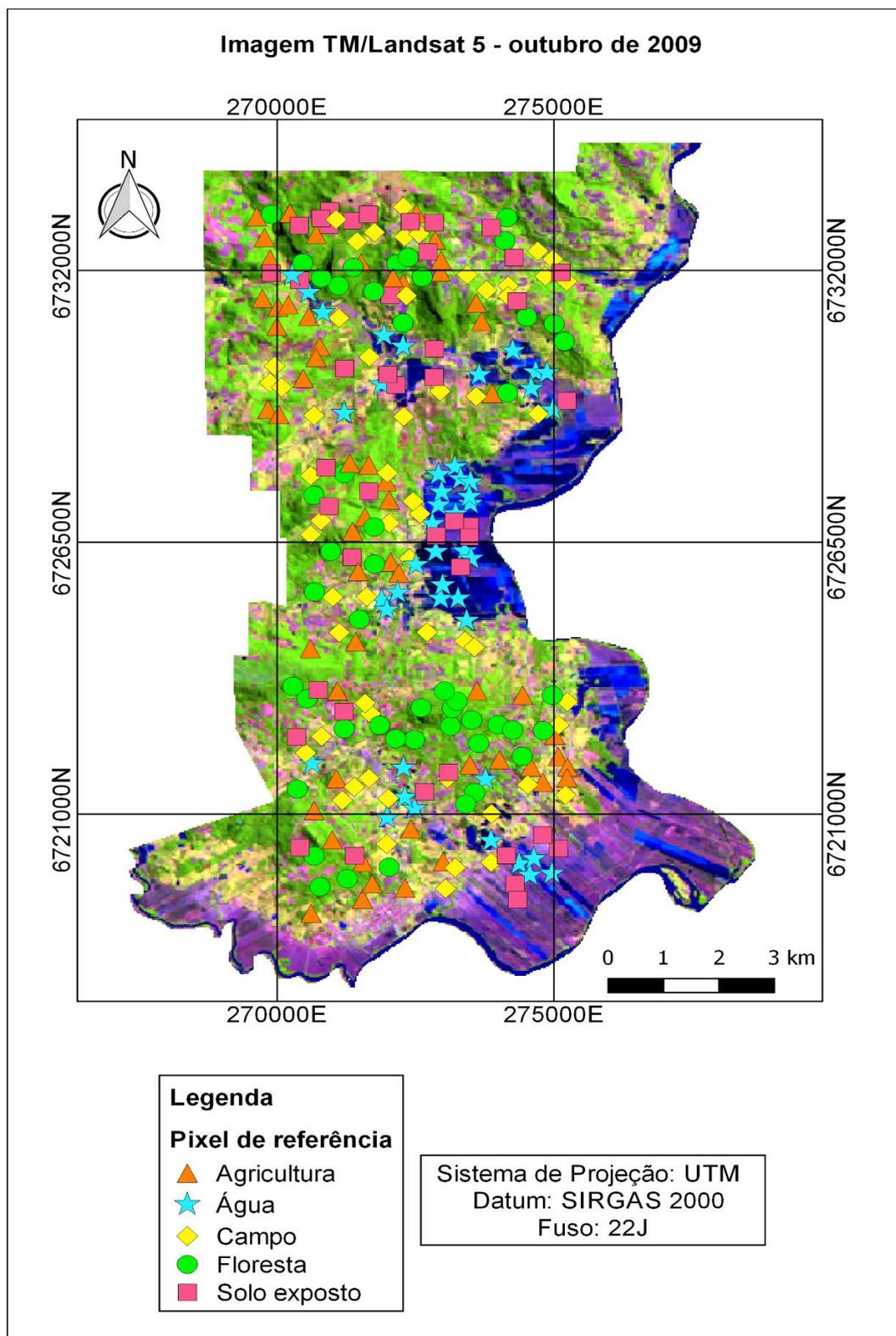
De acordo com o estudo realizado por Faria (2015), no mapeamento das áreas cafeeiras da região das Matas de Minas – MG, mostraram que acima de 80 *pixels* para cada classe, as amostras de treinamento não exerceram influência no desempenho dos classificadores supervisionados como a *Random Forest* e *Support Vector Machine*.

Figura 5 - Identificação das amostras dos *pixels* de referência da imagem sintética RGB 543 do sensor TM/Landsat 5 nos diferentes padrões de uso e cobertura da terra para o ano de 2004, município de Dona Francisca – RS.



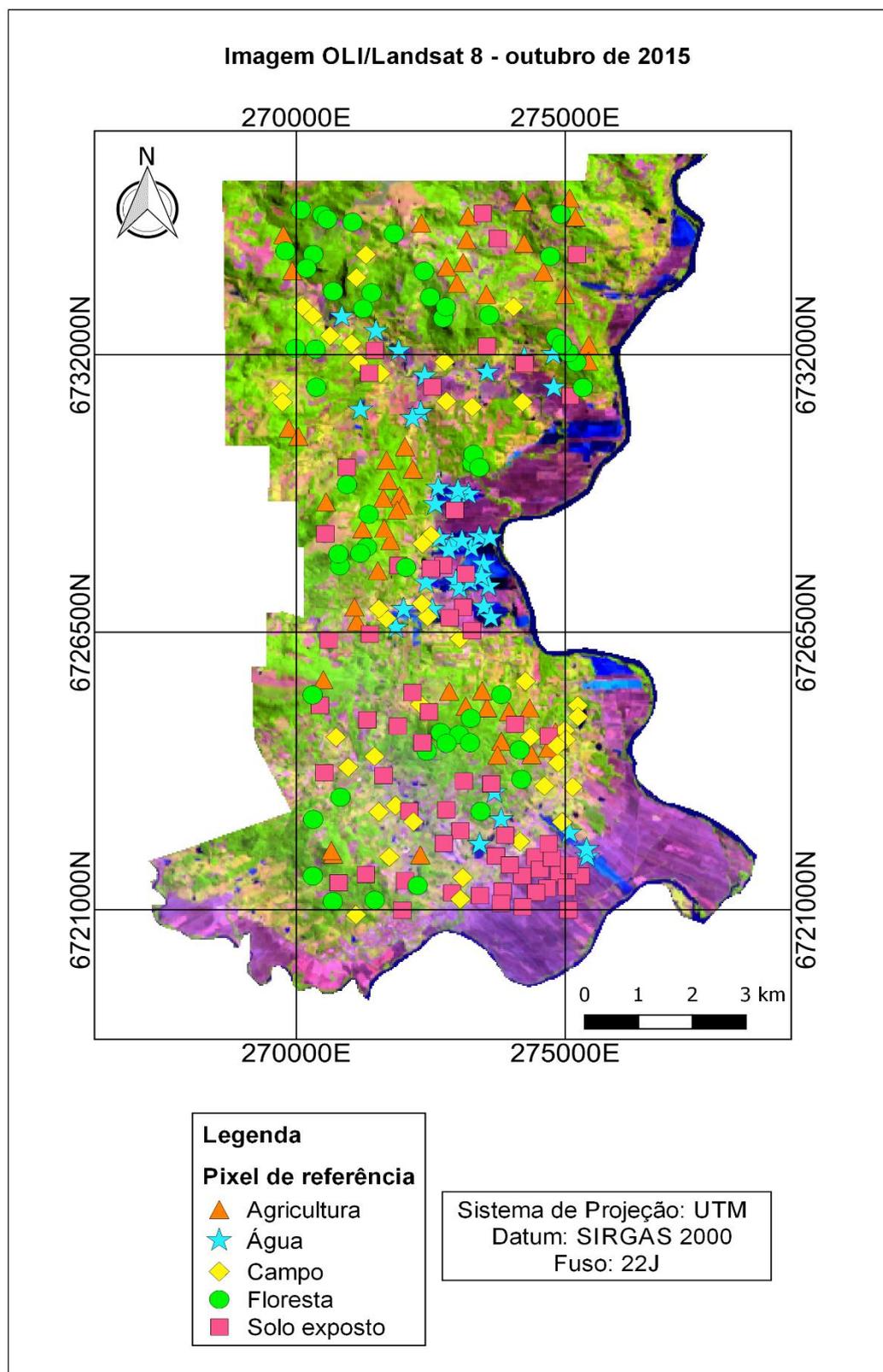
Fonte: Autora (2020).

Figura 6 - Identificação das amostras dos *pixels* de referência da imagem sintética RGB 543 do sensor TM/Landsat 5, nos diferentes padrões de uso e cobertura da terra para o ano de 2009, município de Dona Francisca – RS.



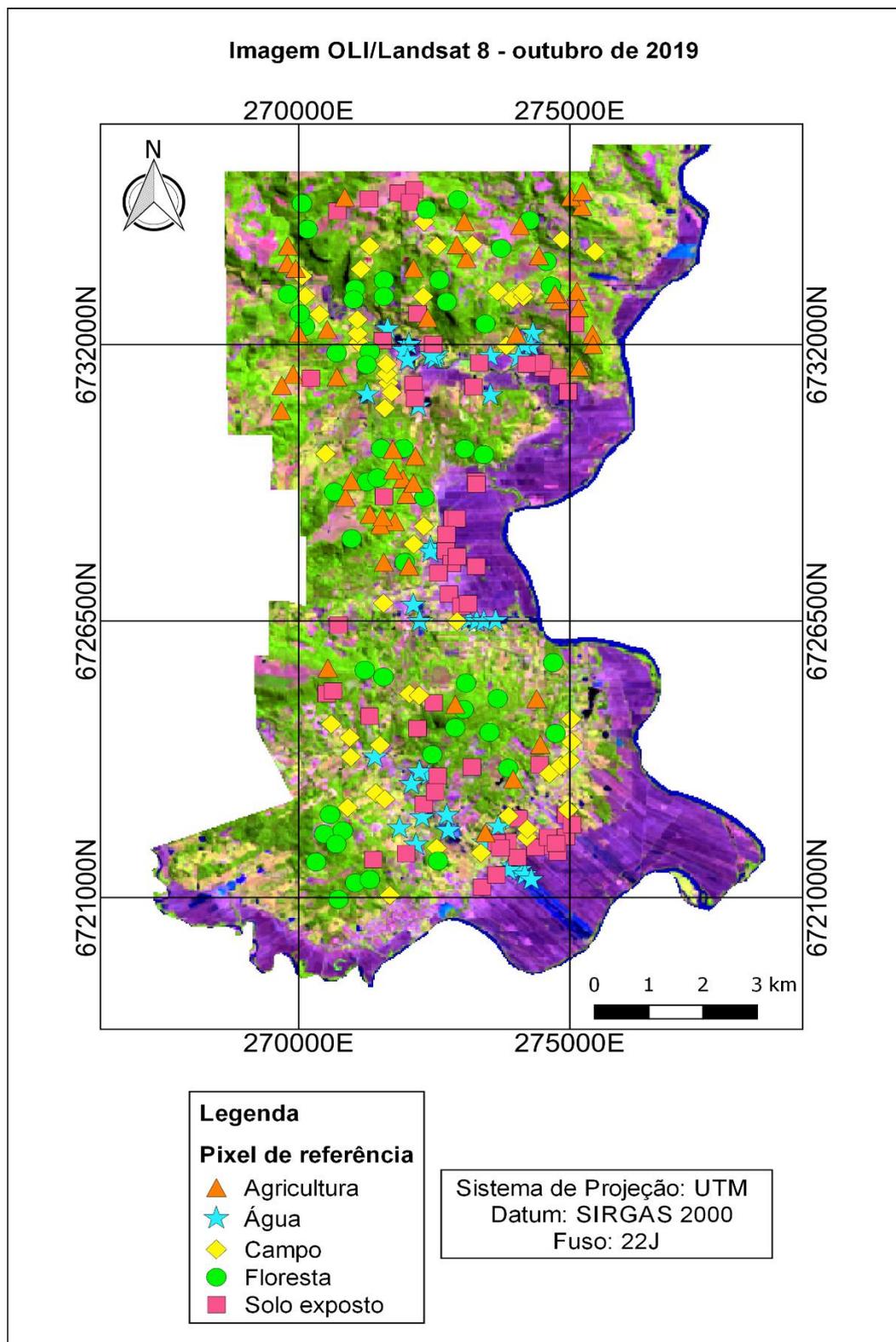
Fonte: Autora (2020).

Figura 7 - Identificação das amostras dos *pixels* de referência da imagem sintética RGB 654 do sensor OLI/Landsat 8, nos diferentes padrões de uso e cobertura da terra para o ano de 2015, município de Dona Francisca – RS.



Fonte: Autora (2020).

Figura 8 - Identificação das amostras dos *pixels* de referência da imagem sintética RGB 654 do sensor OLI/Landsat 8, nos diferentes padrões de uso e cobertura da terra para o ano de 2019, município de Dona Francisca – RS.



4.2 TREINAMENTO, TESTE E DESEMPENHO DOS ALGORITMOS

A etapa da avaliação do desempenho dos algoritmos da *Random Forest* (RF), *K-Nearest Neighbors* (KNN), *Naive Bayes* (NB) e *Support Vector Machine* (SVM) foi realizada por meio das amostras dos *pixels* de referência dos padrões de uso e cobertura da terra das imagens dos sensores TM/Landsat 5 e OLI/Landsat 8, sendo esses dados processado no aprendizado de máquina (treinamento) e avaliados (teste) no *software* KNIME.

4.2.1 Imagem do sensor TM/Landsat 5 – outubro de 2004

Os algoritmos com os melhores desempenhos na predição dos padrões do uso e cobertura da terra e que apresentaram a menor confusão entre esses padrões foram a RF e o SVM, tendo em vista que a acurácia global destes algoritmos para os testes de 20%, 30% e 40% foram 89,5%-91,4%-92% para o RF e para o SVM 87%-88%-90% respectivamente. Os algoritmos RF e SVM segundo Souza (2015), podem alcançar resultados de acurácia global semelhantes.

Usando imagens do satélite Landsat 5 para a classificação da cobertura do solo por meio de algoritmos de AM Santos (2014) averiguou que, os melhores desempenhos foram dos algoritmos SVM com 76,51% e 92,06% e a RF com 77,78% e 90,48%. Moreira et al. (2014), ao utilizarem seis datas provenientes do Landsat 5 para realizar o mapeamento do uso e cobertura do solo da bacia do Rio Vieira – MG obtiveram acurácia do SVM que variaram de 83,51% a 93,23%. Em uma avaliação das imagens do Landsat 5 na classificação de 14 diferentes classes de cobertura da terra no sul da Espanha, Galiano et al. (2012), averiguaram que o algoritmo RF produz classificações precisas de cobertura da terra, com 92%.

Referente aos algoritmos NB e KNN, estes apresentaram qualidade inferior ao desempenho da acurácia (Tabela 6) e, conseqüentemente, maiores inconsistências em predizer os padrões do uso e cobertura da terra. Resultados semelhantes ao *Naive Bayes* do presente estudo foram encontrados por Santos (2014), com desempenho entre 68,25% e 86,03%. Li et al. (2013), usando

imagens do sensor TM/Landsat, em mapeamento de vegetação aquática, obtiveram uma taxa de acerto 86,60% para o NB.

Tabela 6 - Acurácia global do desempenho dos algoritmos na predição dos padrões do uso e cobertura da terra, utilizando o *software* KNIME - outubro de 2004.

15 de outubro de 2004	<i>Random Forest</i>	<i>Support Vector Machine</i>	<i>Naive Bayes</i>	<i>K-Nearest Neighbors</i>	Teste (%)
	89,5%	87%	84,5%	88,5%	20%
	91,4%	88%	86,2%	86,2%	30%
	92%	90%	88%	89,3%	40%

Fonte: Autora (2020).

O índice Kappa dos padrões do uso e cobertura da terra foi fornecido pelo *software* KNIME, representando um índice equivalente para cada algoritmo nos diferentes treinamentos e testes realizados (Tabela 7).

Tabela 7 - Índices de Kappa obtidos pelos algoritmos de aprendizado de máquina, outubro de 2004.

Índice de Kappa - 2004			
Algoritmo	80% - 20%	70% - 30%	60% - 40%
<i>Random Forest</i>	0,86	0,89	0,89
<i>Naive Bayes</i>	0,80	0,82	0,84
<i>Support Vector Machine</i>	0,83	0,85	0,87
<i>K-Nearest Neighbors</i>	0,85	0,82	0,86

Fonte: Autora (2020).

Os resultados obtidos do índice Kappa indicaram que a qualidade das classificações geradas de acordo com Landis e Kock (1977), é de muito boa a excelente variando de 0,80 a 0,89 respectivamente.

4.2.2 Imagem do sensor TM/Landsat 5 – outubro de 2009

A predição dos padrões do uso e cobertura da terra por meio dos *pixels* de referência do Landsat 5 para o ano de 2009, mostrou que a melhor acurácia global foi do algoritmo *Random Forest*, com 90,5%-92,5%-93,3%, seguida do *Support Vector Machine* que também obteve um bom desempenho e bastante semelhante ao RF, com acurácia global de 91,5%-90,8%-93,3%, demonstrando assim, uma classificação adequada dos algoritmos de aprendizado de máquina quanto ao número de *pixels* pertencentes a cada padrão do uso e cobertura da terra.

Para a classificação ser considerada adequada, segundo os autores Hentz et al. (2015); Pereira e Santos (2016) busca-se que seus valores sejam altos, ou seja, quanto maior for o valor da exatidão global, maior será a quantidade de *pixels* classificados corretamente.

Para Pal (2005), os algoritmos RF e SVM podem alcançar resultados de precisões de classificação semelhantes, resultados esses que podem ser confirmados com os dados obtidos neste estudo. De acordo com Mountrakis et al. (2011), o SVM tem sido amplamente usado, pois apresenta superioridade sobre os demais e uma boa capacidade de generalização mesmo com limitado número de amostras. Nessa mesma linha, a principal característica do SVM, para Niknejad et al. (2014), é a alta capacidade de usar menos amostras de treinamento para atingir níveis mais altos de precisão em comparação com outros métodos.

Utilizando dados das imagens de satélite do Landsat 5 para análise espaço temporal do uso e ocupação do solo na bacia do Rio Vieira – MG, Moreira et al. (2014), constataram que no mapeamento de 2010 o algoritmo SVM obteve a melhor acurácia com 93,23%, devido à melhor separabilidade espectral das amostras de referência. Pereira e Centeno (2013) e Sousa et al. (2010), em seus estudos obtiveram resultados satisfatórios ao empregar o algoritmo SVM na discriminação de classes de uso e ocupação do solo, com elevado grau de acurácia, correspondendo a 99,6%, seguido de 86,03% respectivamente,

Estudos realizados por Tian et al. (2016), para classificar a cobertura de terra úmidas na bacia do rio Ertix na China, por meio de imagens do Landsat 8, averiguaram que no treinamento de 80% e 20% de teste, a RF apresentou

acurácia de 92,5%, sendo essa satisfatória. Para Wieland e Pittore (2014); Li et al. (2014); Duro et al. (2012); Gislason et al. (2006), os algoritmos *Random Forest* e *Support Vector Machine* apresentam resultados promissores para a classificação em diferentes sensores, enquanto que o algoritmo NB apresentou os piores resultados. Corroborando os dados obtidos neste estudo, no qual o *Naive Bayes* apresentou novamente acurácia inferior com 82,5%-84,5%-88% quando comparado com os demais algoritmos. Já para o *K-Nearest Neighbors* a quantidade de *pixels* classificados corretamente foi de 88%-89,1%-90,5%, para os testes de 20%, 30% e 40% respectivamente (Tabela 8).

Tabela 8 - Acurácia global do desempenho dos algoritmos nas predições dos padrões do uso e cobertura da terra, utilizando o *software* KNIME - outubro de 2009.

13 de outubro de 2009	<i>Random Forest</i>	<i>Support Vector Machine</i>	<i>Naive Bayes</i>	<i>K-Nearest Neighbors</i>	Teste (%)
	90,5%	91,5%	82,5%	88%	20%
	92,5%	90,8%	84,5%	89,1%	30%
	93,3%	93,3%	86%	90,5%	40%

Fonte: Autora (2020).

A classificação proposta por Landis e Kock (1977), descrita na Tabela 9, permite afirmar que a classificação variou de muito boa a excelente, uma vez que o índice de Kappa encontrado está presente na faixa entre 0,78-0,91.

Tabela 9 - Índices de Kappa obtidos pelos algoritmos de aprendizado de máquina, outubro de 2009.

Índice de Kappa - 2009			
Algoritmo	80% - 20%	70% - 30%	60% - 40%
<i>Random Forest</i>	0,88	0,90	0,91
<i>Naive Bayes</i>	0,78	0,80	0,82
<i>Support Vector Machine</i>	0,89	0,88	0,91
<i>K-Nearest Neighbors</i>	0,84	0,86	0,88

Fonte: Autora (2020).

4.2.3 Imagem do sensor OLI/Landsat 8 – outubro de 2015

A predição dos padrões do uso e cobertura da terra do Landsat 8 para o ano de 2015, quando processados no *software* KNIME, demonstraram que os algoritmos *K-Nearest Neighbors* e *Random Forest* apresentaram os melhores desempenhos, no qual previu uma acurácia global de 91,5%-92%-94% e 91%-93,7%-90,6%, respectivamente. Estes valores são superiores aos resultados relatados na classificação do uso da terra no sul do Peru por TATSUMI et al., (2015), a qual obteve uma precisão geral de 81% para o RF apontando como um resultado de alta performance.

Müller et al. (2015) analisando o algoritmo RF em séries temporais de Landsat 8 para classificar culturas, pastagem e vegetação natural obtiveram exatidão global de até 93%. Avaliando diversos classificadores para o mapeamento de espécies lenhosas invasoras, nos Açores em Portugal, Gil et al. (2014), encontraram os melhores resultados de acurácia para o KNN variando de 89% a 94%. Sarmiento et al. (2014) mostraram que o algoritmo KNN apresentou bons índices (90% de acerto) para o mapeamento de áreas cafeeiras em Campos Gerais, no Estado de Minas Gerais.

O algoritmo *Naive Bayes*, assim como o *Support Vector Machine* apresentaram os menores índices de desempenho, com 78%-79%-84,6% e 83,5%-88%-90% respectivamente (Tabela 10). Apesar desses valores consideravelmente mais baixos em relação aos outros algoritmos, o desempenho destes pode ser considerado satisfatório.

Tabela 10 - Acurácia global do desempenho dos algoritmos nas predições dos padrões do uso e cobertura da terra, utilizando o *software* KNIME - outubro de 2015.

05 de outubro de 2015	<i>Random Forest</i>	<i>Support Vector Machine</i>	<i>Naive Bayes</i>	<i>K-Nearest Neighbors</i>	Teste (%)
	91%	83,5%	78%	91,5%	20%
	93,7%	88%	79%	92%	30%
	90,6%	90%	84,6%	94%	40%

Fonte: Autora (2020).

Os resultados obtidos neste estudo para o SVM estão em concordância com os encontrados por Souza et al. (2016), que mesmo trabalhando com o mapeamento da cafeicultura nos municípios de Araguari, Carmo de Minas, Três Pontas em Minas Gerais, averiguaram que as melhores acurácias foram do SVM, com 85,33%, 87% e 88,33% respectivamente. Otukey e Blaschke (2010), comparando os algoritmos de classificação *Support Vector Machines*, *Árvore de Decisão* e *Máxima Verossimilhança*, nas mudanças da cobertura do solo por meio de imagens Landsat 7 ETM+ de dois diferentes anos, tiveram como resultado a acurácia de mais de 85% de acerto.

Dhima et al. (2017), usando também o *software* KNIME para análise dos dados de AM de diferentes tipos de solo, constataram que o SVM apresentou uma acurácia de aproximadamente 85%, enquanto o NB obteve uma taxa de acerto de 75%. Albuquerque (2012), utilizando o algoritmo SVM para analisar a cobertura do solo no entorno a Usina de Lajeado – TO, obteve uma acurácia de 86,11%.

Usando o algoritmo *Naive Bayes*, na classificação de 10 espécies provenientes do banco de dados de folhas da Mata Atlântica, Souza e Kai (2014), tiveram como resposta uma taxa de acerto de 70,56%. Assim como os resultados de Cracknell e Reading (2014), mostraram que o algoritmo NB não obteve os melhores resultados de acurácia, variando de 0,54% até 0,61% usando imagens Landsat 7 ETM+ para mapear áreas geológicas.

Em relação ao índice Kappa, conforme a classificação definida por Landis e Koch (1977) no presente estudo corresponde a uma classificação muito boa a excelente (Tabela 11).

Tabela 11 - Índices de Kappa obtidos pelos algoritmos de aprendizado de máquina, outubro de 2015.

Índice de Kappa - 2015			
Algoritmo	80% - 20%	70% - 30%	60% - 40%
<i>Random Forest</i>	0,88	0,92	0,88
<i>Naive Bayes</i>	0,72	0,74	0,80
<i>Support Vector Machine</i>	0,79	0,85	0,87
<i>K-Nearest Neighbors</i>	0,89	0,90	0,92

Fonte: Autora (2020).

4.2.4 Imagem do sensor OLI/Landsat 8 – outubro de 2019

A predição dos padrões do uso e cobertura da terra para o ano de 2019, por meio do aprendizado de máquina, demonstrou que o algoritmo *K-Nearest Neighbors* apresentou os melhores desempenhos com 90,5%-91%-92%, seguido da *Random Forest* com 88%-90,8%-91,3%. Os resultados da classificação da cobertura da terra de zonas úmidas costeiras por Wang et al. (2019), demonstraram que as precisões da RF, SVM e KNN foram 86,61%, 79,96% e 77,23%, respectivamente, para imagens do Landsat 8. Jin et al. (2011), utilizando dados do satélite multiespectral de alta resolução QuickBird para classificar a cobertura do solo de uma região de Xian na China, constataram que a precisão por meio do algoritmo KNN foi de até 91,58%.

Ao observarmos os valores do desempenho dos algoritmos *Naive Bayes*, com 77%-80%-81,3% e o *Support Vector Machine*, com 84%-81%-86% nota-se que ambos apresentaram resultados inferiores quando comparado com os demais (Tabela 12). Ainda assim, resultados obtidos do índice Kappa com 0,70 para o algoritmo NB comprovam a eficiência da classificação sendo definida por Landis e Kock (1977), como muito boa (Tabela 13).

Topaloğlu et al. (2016), ao avaliarem a precisão usando o algoritmo SVM em imagens do Landsat 8, encontraram uma acurácia de 81,67% com um índice de Kappa de 0,79. Bulut e Günlü (2016), ao classificar as classes de uso da terra usando o Landsat 8, averiguaram que o SVM obteve uma precisão geral de 84% e um índice Kappa de 0,79.

Tabela 12 - Acurácia global do desempenho dos algoritmos nas predições dos padrões do uso e cobertura da terra, utilizando o *software* KNIME - outubro de 2019.

09 de outubro de 2019	<i>Random Forest</i>	<i>Support Vector Machine</i>	<i>Naive Bayes</i>	<i>K-Nearest Neighbors</i>	Teste (%)
	88%	84%	77%	90,5%	20%
	90,8%	81%	80%	91%	30%
	91,3%	86%	81,3%	92%	40%

Fonte: Autora (2020).

Tabela 13 - Índices de Kappa obtidos pelos algoritmos de aprendizado de máquina, outubro de 2019.

Índice de Kappa - 2019			
Algoritmo	80% - 20%	70% - 30%	60% - 40%
<i>Random Forest</i>	0,84	0,88	0,89
<i>Naive Bayes</i>	0,70	0,74	0,76
<i>Support Vector Machine</i>	0,79	0,75	0,82
<i>K-Nearest Neighbors</i>	0,88	0,89	0,90

Fonte: Autora (2020).

Analisando de forma geral, os algoritmos mostraram desempenhos diferentes, com superioridade nos valores do índice de Kappa para a *Random Forest* para os anos de 2004 e 2009, enquanto que o *K-Nearest Neighbors* apresentou essa superioridade nos anos 2015 e 2019.

Percebeu-se ainda, que as amostras dos *pixels* de referência dos sensores TM/Landsat 5 e OLI/Landsat 8, aliadas aos algoritmos de AM podem auxiliar de forma positiva, conseguindo distinguir adequadamente os diferentes padrões de uso e cobertura da terra.

5 CONSIDERAÇÕES FINAIS

Os algoritmos com os melhores desempenhos na classificação dos *pixels* de referência dos padrões de uso e cobertura da terra e que apresentaram a menor confusão foram a *Random Forest* e *Support Vector Machine* para os anos de 2004 e 2009. Com relação à análise do desempenho entre os algoritmos para os valores dos *pixels* de referência das imagens de 2015 e 2019, o algoritmo *K-Nearest Neighbors* e a *Random Forest* tiveram os resultados mais altos para as acurácias globais do que os algoritmos *Support Vector Machine* e *Naive Bayes*.

Quando comparados aos demais algoritmos a *Random Forest*, *K-Nearest Neighbors* e *Support Vector Machine*, o algoritmo *Naive Bayes* apresentou desempenhos inferiores em todas as datas estudadas, que variaram de 77% a 88% de acurácia global.

Os valores do índice Kappa gerados pelo *software* KNIME indicam que a qualidade das classificações geradas pelos algoritmos RF, SVM, KNN e NB para os anos 2004, 2009, 2015 e 2019 variaram de muito boa a excelente.

Os dados provenientes das amostras dos *pixels* de referência dos padrões do uso e cobertura da terra das imagens dos sensores TM/Landsat 5 e OLI/Landsat 8, demonstraram ser eficientes para o processo do aprendizado de máquina. Estes algoritmos tiveram resultados satisfatórios, de maneira que apresentaram eficiência na classificação dos padrões do uso e cobertura da terra a partir de dados provenientes das imagens orbitais.

Como contribuição deste trabalho, cito a análise de desempenho dos algoritmos de aprendizado de máquina, processados na plataforma de análise de dados KNIME. Uma vez que, a disseminação de informações relativas a essa plataforma ainda é pouca ou inexistente em grandes periódicos e base de dados na área da Engenharia Florestal, principalmente no que diz respeito ao aprendizado de máquina utilizando séries temporais providas dos sensores remotos.

Cabe ainda ressaltar que as plataformas *open source* como o QGIS e KNIME, bem como as imagens do satélite Landsat 5 e Landsat 8 são importantes para a comunidade como um todo, pois suas licenças são disponibilizadas de forma gratuita. Estes *softwares* possibilitam aos usuários a liberdade de estudar a dinâmica de funcionamento dos seus respectivos programas, bem como,

adaptá-los para as suas necessidades de modo que toda a comunidade de profissionais, os que utilizam como ferramentas para as suas práticas diárias ou para estudos em pesquisas científicas, possa ser beneficiada.

Além disso as imagens geradas em diferentes resoluções, diferentes bandas espectrais e de diversas épocas, permitem que as instituições de pesquisa, setores privados e governamentais tenham uma maior capacidade de realizar análises sobre os padrões de uso e cobertura da terra em diversas escalas espaço-temporais.

REFERÊNCIAS

- ADAM, E.; MUTANGA, O.; ODINDI, J.; ABDEL-RAHMAN, E. M. Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. **International Journal of Remote Sensing**, v. 35, n. 10, p. 3440-3458, 2014.
- ALBUQUERQUE, R. W. DE. 2012. 95 p. **Monitoramento da cobertura do solo no entorno de hidrelétricas utilizando o classificador SVM (Support Vector Machines)**. Dissertação (Mestrado em Engenharia de Transporte) - Universidade de São Paulo, São Paulo, 2012.
- ALI, I.; GREIFENEDER, F.; STAMENKOVIC, J.; NEUMANN, M.; NOTARNICOLA, C. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. **Remote Sensing**, v. 7, n. 12, p. 16398-16421, 2015.
- ALPAYDIN, E. **Introduction to machine learning**. MIT Press, 2010.
- ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONÇALVES, J. L. M.; SPAROVEK G. Köppen's climate classification map for Brazil. **Meteorologisch e Zeitschrift**, Berlin, v. 22, n. 6, p. 711-728. 2013.
- AMARAL, F. **Aprenda Mineração de Dados: Teoria e prática**. São Paulo: Alta Books, 2016.
- ANDRADE, C. S. M. de. 2015. 119 p. **Text Mining na Análise de Sentimentos em Contextos de Big Data**. Dissertação (Mestrado em Engenharia e Gestão de Sistemas de Informação) – Universidade do Minho, Braga, 2015.
- ANTUNES, R. R. 2018. 45 p. **Integração, controle e acompanhamento da análise de imagens baseada em objeto e mineração de dados por meio da plataforma distribuída intercloud**. Tese (Doutorado em Geociências) - Instituto de Geociências – IG, Brasília, Distrito Federal, 2018.
- BASHA, S. M.; RAJPUT, D. S., POLURU, R. K.; BHUSHAN, S. B.; BASHA, S. A. K. Evaluating the performance of supervised classification models: decision tree and Naïve Bayes using KNIME. **International Journal of Engineering & Technology**, v. 7, n. 4.5, p. 248-253, 2018.
- BELGIU, M.; DRĂGU, L. Random forest in remote sensing: A review of applications and future directions. **Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.
- BENEDETTI, A. C. P. 2010. 167 p. **Modelagem dinâmica para simulação de mudanças na cobertura florestal das Serras do Sudeste e Campanha**

Meridional do Rio Grande do Sul. Tese (Doutorado em Engenharia Florestal) - Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, 2010.

BERTHOLD, M. R.; CEBRON, N.; DILL, F.; GABRIEL, T. R.; KÖTTER, T.; MEINL, T.; OHL, P.; SIEB, C.; THIEL, K.; WISWEDEL, B. **KNIME: The Konstanz Information Miner.** In: Preisach C., Burkhardt H., Schmidt-Thieme L., Decker R. (eds) Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, Heidelberg. 2008.

BISCHOF, H.; SCHNEIDER, W.; PINZ, A. J. Multispectral classification of LANDSAT-images using neurais networks. **Transactions Geoscience and Remote Sensing**, v. 30, n. 3, p. 482-490, 1992.

BORGES, L. M C. **Um estudo da Diabetes Mellitus e Hipertensão Arterial baseado em técnicas de Data Mining aplicadas a dados da Administração Regional de Saúde do Centro.** 2016. 184 p. Dissertação (Mestrado em Informática e Sistemas). Instituto Politécnico de Coimbra, Coimbra, Portugal, 2016.

BRAGA, R. de. C. 2015. 138 p. **Estratégias Integradas em Química Medicinal para a identificação de novos compostos bioativos contra *Leishmania infantum*.** Tese (Doutorado em Química) - Universidade Federal de Goiás, Goiânia, 2015.

BREIMAN, L. **Random Forests.** **Machine Learning**, v. 45, n. 1, p. 5-32, 2001.

BRITES, R. S.; SOARES, V. P.; RIBEIRO, C. A. A. Comparação de Desempenho entre Três Índices de Exatidão Aplicados a Classificações de Imagens Orbitais. VIII Simpósio Brasileiro de Sensoriamento Remoto. **Anais...** Salvador. 1996. p. 813-821.

BUDUMA, N; LOCASCIO, N. **Fundamentals of deep learning: Designing next-generation machine intelligence algorithms.** " O'Reilly Media, Inc.", 2017. 298 p.

BULUT, S.; GÜNLÜ, A. Arazi kullanım sınıfları için farklı kontrollü sınıflandırma algoritmalarının karşılaştırılması. **Kastamonu Üniversitesi, Orman Fakültesi Dergisi**, v. 16, n. 2, p. 528-535, 2016.

BUNTING, P.; CLEWLEY, D.; LUCAS, R.M. E GILLINGHAM, S. The Remote Sensing and GIS Software Library (RSGISLib), **Computers and Geosciences**, v. 62, p. 216-226, 2014.

CALIJURI, M. do C.; CUNHA, D. G. F. Engenharia Ambiental: conceitos, tecnologia e gestão. **Edição Única. Rio de Janeiro: Elsevier**, 2013.

CÂMARA, G.; MONTEIRO, A. M. V.; DE MEDEIROS, J. S. Representações Computacionais do Espaço: Fundamentos Epistemológicos da Ciência da Geoinformação. **Geografia**, Rio Claro, v. 28, n. 1, p. 83-96, 2003.

CÂMARA, G.; QUEIROZ, G. R. Arquitetura de sistema de informação geográfica. In: CÂMARA, G. MONTEIRO, A. M. V.; MEDEIROS, J. S. **Introdução à Ciência da Geoinformação**. São José dos Campos, SP: INPE, 2004.

CARVALHO, A. **Inteligência Artificial - Uma Abordagem de Aprendizado de Máquina**. 1ª Ed. São Paulo: LTC, 2011. 394 p.

CARVALHO, S. S. de. M. 2017. 79 p. **Previsão da duração de procedimentos cirúrgicos num hospital**. Dissertação (Mestrado em Modelação Análise de Dados e Sistemas de Apoio à Decisão) - Universidade do Porto, Portugal, 2017.

CARVALHO, I. S. de.; FERRAZ, T. M.; SILVA, T. S. F.; ALVARADO, S. T. Classificação da vegetação do Parque Nacional da Chapada das Mesas, Maranhão, usando OBIA, machine learning e softwares livres. XIX do Simpósio Brasileiro de Sensoriamento Remoto. **Anais...**Santos. 2019. p. 2984-2987.

CAVALCANTI, L. C. A. M. 2016. 94 p. **Detecção de elementos antrópicos e imagens aéreas da floresta Amazônica**. Dissertação (Mestrado em Informática). Universidade Federal de Amazonas, Manaus, 2016.

CAVALCANTI, E. B. V. S. 2018. 298 p. **Estudos quimiotaxonômicos e triagem virtual de flavonoides isolados da família Asteraceae com potencial atividade leishmanicida**. Tese (Doutorado em Produtos Naturais e Sintéticos Bioativos) – Universidade da Paraíba, João Pessoa, 2018.

CECCONELLO, S. T. 2017. 98 p. **Análise ambiental dos processos dinâmicos do uso e cobertura da terra sobre as áreas de proteção permanente no município de Pelotas entre 1985 e 2015**. Dissertação (Mestrado em Desenvolvimento Territorial e Sistemas Agroindustriais), Universidade Federal de Pelotas, Pelotas, Rio Grande do Sul, 2017.

CHEN, L.; SYCARA, K. **WebMate: A personal agente for browsing and searching**. Pittsburgh, EUA: Carnegie Mellon University, v. 98, p. 132-139, 1997.

CONGALTON, R. G. A Review of assessing the accuracy of classifications of remotely sensed data. **Remote Sensing Environment**, v. 37, p. 35-46, 1991

CONGALTON, R. G.; BIGING, G. S. A pilot study evaluating ground reference data collection efforts for use in forest inventory. **Photogrammetric Engineering and Remote Sensing**, v. 58, n. 12, p. 1669-1671, 1992.

CONGALTON, R. G.; GREEN, K. A practical look at source of confusion in error matrix generation. **Photogrammetric engineering and remote sensing (USA)**, v. 59, n. 5, p. 641-644, 1993.

- CONGEDO, L. 2019. **Tutorial: Land Cover Classification Using the Semiautomatic Classification Plugin v. 3.0 “Rome” for QGIS**. Disponível em: <<http://fromgistors.blogspot.com/>>. Acesso em: 10 de jan. 2019.
- CORTES, C.; V. VAPNIK. 1995. Support-Vector Networks. **Machine Learning**, v. 20, n. 3, p. 273-297, 1995.
- COSME, A. **Projeto em Sistemas de Informação Geográfica**. Lisboa: Ed. Lidel, 2012.
- CPRM. **Serviço Geológico do Brasil**. Disponível em: <<http://geobank.cprm.gov.br/>>. Acesso em: 10 de jan. 2019.
- CRACKNELL, M. J.; READING, A. M. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. **Computers & Geosciences**, New York, v. 63, p. 22-33, 2014.
- CRISTIANINI, N.; SHAW-TAYLOR, J. **An introduction to support vector machines and other kernel based learning methods**. Cambridge: Cambridge University, 2001. 191p.
- CRÓSTA, A. P. **Processamento Digital de Imagens de Sensoriamento Remoto**. Campinas: IG/UNICAMP, 1992. 170 p.
- CUTLER, R., EDWARDS, T. C., BEARD, K. H. CUTLER, A., HESS, K. T., GIBSON, J., LAWLER, J. J. Random Forests For Classification in ecology. **Ecology**, v. 88, n. 11, p. 2783-2792, 2007.
- DHIMAN, V.; BEDI, R.; KUMAR, M. Predictive big data analysis for machine learning using Knime to solve many business challenges. International Interdisciplinary Conference on Science Technology Engineering Management Pharmacy and Humanities. **Anais...** Singapore. 2017. p. 329-335.
- DOMINGOS, P.; PAZZANI, M. Simple Bayesian classifiers do not assume independence, In: **Proceedings of the National Conference on Artificial Intelligence**, p. 1386, 1996.
- DUDA, R.; HART, P.; STORK, D. **Pattern classification and scene analysis**. 2. ed. New York: John Wiley & Sons, 2000.
- DURO, D. C.; FRANKLIN, S. E.; DUBÉ, M. G. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. **Remote Sensing of Environment**, v. 118, p. 259-272, 2012.
- EMBRAPA. **Monitoramento por satélite**. Satélites de Monitoramento. Campinas: Embrapa Monitoramento por Satélite. 2013. Disponível em: <<http://www.sat.cnpm.embrapa.br>>. Acesso em: 11 abr. 2019.

ESTEVEES, S. R.; LORENA, A. C.; NASCIMENTO, M. Z. **Aplicação de técnicas aprendizado de máquina na classificação de imagens mamográficas**. In: Simpósio de iniciação científica da Universidade Federal do ABC, p. 1-2. 2009.

FARIA, M. M. 2015. 89 p. **Distribuição espacial do café na região das matas de Minas**. Tese (Doutorado em Solos e Nutrição de Plantas) - Universidade Federal de Viçosa, Viçosa, Minas Gerais, 2015.

FASSNACHT, F. E.; NEUMANN, C.; FÖRSTER, M.; BUDDENBAUM, H.; GHOSH, A.; CLASEN, A.; JOSHI, P. K.; KOCH, B. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. **Revista IEEE de Tópicos Seleccionados em Observações Aplicadas à Terra e Sensoriamento Remoto**, v. 7, n. 6, p. 2547-2561, 2014.

FAVERA, A. D. 2006. 127 p. **O cenário produtivo e a representatividade dos rios Jacuí e Soturno para o município de Dona Francisca – RS**. Dissertação (Mestrado em Geografia). Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, 2006.

FELTRIN, L. KNIME an Open Source Solution for Predictive Analytics in the Geosciences [Software and Data Sets]. **IEEE Geoscience and Remote Sensing Magazine**, v. 3, n. 4, p. 28-38, 2015.

FLORENZANO, T. G. **Iniciação em Sensoriamento Remoto**. São Paulo: Oficina de Textos, 2007.

FLORENZANO, T. G. **Iniciação em Sensoriamento Remoto**. 3ª ed. São Paulo: Oficina de Textos, 2011. 128 p.

FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. Bayesian Network Classifiers. **Machine learning**, v. 29, n. 2-3, p. 131-163, 1997.

FRIEDRICH, C. 2014. 181 p. **Comparison of ArcGIS and QGIS for Applications in Sustainable Spatial Planning**. Dissertação (Mestrado em Ciências), Universidade de Viena, 2014.

GANEM, K. A. 2017. 202 p. **Classificação da Cobertura do Solo na Caatinga a partir de Imagens do Landsat-8 e da Ferramenta Google Earth Engine: uma comparação entre dados com e sem correção atmosférica**. Dissertação (Mestrado em Geociências Aplicadas) - Universidade de Brasília, Distrito Federal, 2017.

GALIANO, R. V. F.; GHIMIRE, B.; ROGAN, J.; CHICA-OLMO, M.; RIGOL SANCHEZ, J. P. An assessment of the effectiveness of a random forest classifier for land-cover classification. **ISPRS Journal Photogrammetry & Remote Sensing**, Calgary-AB, v. 67, p. 93–104, 2012.

GAO Y.; GAO, F. Edited AdaBoost by weighted KNN. **Neurocomputing**, v. 73, n. 16-18, p. 3079-3088, 2010.

GELETE, C.; VOLATÃO, C. F. de S. Análise do potencial de classificação do uso e cobertura do solo por meio de rede neural. XIII Simpósio Brasileiro de Sensoriamento Remoto. **Anais...** Florianópolis. 2007. p. 5779-5785.

GHOSH, A.; FASSNACHT, F. E.; JOSHI, P. K.; KOCH, B. A framework for mapping tree species combining hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales. **International Journal of Applied Earth Observation and Geoinformation**, v. 26, p.49-63, 2014.

GOLDSCHIMIDT, R. R. **Inteligência Computacional**. 1ª ed. Rio de Janeiro: Instituto Superior de Tecnologia - Rio, 2010. 143 p.

GIL, A.; YU, Q.; ABADI, M.; CALADO, H. Using aster multispectral imagery for mapping woodyinvasive species in Pico da Vara natural reserve (Azores Islands, Portugal). **Revista Árvore**, Viçosa, v. 38, n. 3, p. 391-401, 2014.

GISLASON, P. O.; BENEDIKTSSON, J. A.; SVEINSSON, J. R. Random Forests for land cover classification. **Pattern Recognition Letters**, v. 27, n. 4, p. 294-300, 2006.

GONÇALVES, P.; CARRÃO, H.; PINHEIRO, A; CAETANO, M. Land cover classification with Support Vector Machine Applied to MODIS imagery. In: GLOBAL DEVELOPMENTS IN ENVIRONMENTAL EARTH OBSERVATION FROM SPACE, 25., 2006, Porto. **Proccedings...**Porto: Millpress Science Publishers, 2006. p. 517-526.

HARTMANN, C.; BULLA, L. A. S.; FELLINI B. D. Emprego de Mosaico de médio formato e do Geoprocessamento na definição dos limites técnicos para extração da areia do baixo rio Jacuí-RS, Brasil. **Gravel**, v. 8, n. 1. p. 9-19. 2010.

HENRIQUES, A. de. P. de M. 2008. 148 p. **Classificação de imagens de ambientes coralinos: uma abordagem empregando uma combinação de classificadores e Máquina de Vetor de Suporte**. Tese (Doutorado em Engenharia Elétrica e de Computação) – Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, 2008.

HENTZ, Â. M. K.; Dalla CORTE, A. P.; BLUM, C. T.; SANQUETTA, C. R. Técnicas orientadas ao objeto para levantamento da fragmentação florestal na sub-bacia Alto Iguaçu, Paraná. **Geociências**, São Paulo, v. 34, n. 4, p.883-896, 2015.

IBGE. **Instituto Brasileiro de Geografia e Estatística**. Disponível em: <www.cidades.ibge.gov.br>. Acesso em: 10 de jan. 2019.

INPE. **Instituto Nacional de Pesquisas Espaciais**. 2019. Disponível em: <http://www.dpi.inpe.br/spring/portugues/tutorial/introducao_geo.html>. Acesso em: 10 de jan. 2019.

IRONS, J. R.; DWYER, J. L.; BARSİ, J. A. The next Landsat satellite: The Landsat data continuity mission. **Remote Sensing of Environment**, v. 122, p. 11-21, 2012.

IWAI, O. K.; QUINTANILHA, J. A. Utilização de imagens de satélite como ferramenta de auxílio ao planejamento urbano. **Revista de Geografia Norte Grande**, n. 34, p. 65-82, 2005.

JENSEN, J. R. **Sensoriamento remoto do ambiente uma perspectiva em recursos terrestres**. Tradução de José Carlos Neves Epiphany et al., São José dos Campos: Editora Parêntese, 2009. 598 p.

JIN, Z.; PU, Y.; MA, J.; CHEN, G. The geographical weighted K-NN classifiers in land cover classification from remote sensing image: A case study of a subregion of Xi'an, China. In: **19th International Conference on Geoinformatics**. IEEE, 2011. p. 1-5.

JÚNIOR, A. M. 2018. 74 p. **Geoestatística na avaliação do teor de clorofila na aveia preta**. Tese (Doutorado em Agronomia) - Universidade Estadual Paulista, Faculdade de Ciências Agrônomicas, Botucatu, São Paulo, 2018.

KAWASHIMA, R. S. 2015. 190 p. **Modelagem da dinâmica espacial das mudanças na cobertura da terra na região portuária da Baixada Santista**. Dissertação (Mestrado em Ciências) - Escola Politécnica da Universidade de São Paulo, São Paulo, 2015.

KNIME. **Konstanz Information Miner**. 2019. Disponível em: <<http://www.knime.org/>>. Acesso em: 10 de jan. 2019.

KNIME. **Konstanz Information Miner**. 2019. Disponível em: <<https://www.knime.com/blog/what-about-data-access>>. Acesso em: 10 de jan. 2019.

LAGO, R. B.; FARENZENA, D. Contrastes socioeconômicos do conjunto habitacional Gentil Tessele e área central da cidade de Dona Francisca, RS. **Disciplinaram Scientia**. Série: Ciências Humanas, Santa Maria, v. 9, n. 1, p. 17-34. 2008.

LANDIS, J.; KOCH, G. G. The measurements of agreement for categorical data. **Biometrics**, Washington, v. 33, n. 3, p. 159-179, 1977.

LARY, D. J.; AMIR, H. A.; AMIR, H. G.; CANNETTE, L. Machine learning in geosciences and remote sensing. **Geoscience Frontiers**, v. 7, n.1, p. 3-10. 2016.

LI, F.; LI, C.; XIAO, B.; WANG, Y. Mapping large-scale distribution and changes of aquatic vegetation in Honghu Lake, China, using multitemporal satellite imagery. **Journal of Applied Remote Sensing**, v. 7, p. 1-16, 2013.

LI, C.; WANG, J.; WANG, L.; HU, L.; GONG, P. Comparison of classification algorithms and training sample sizes in urban land classification with Landsat Thematic Mapper imagery. **Remote Sensing**, v. 6, n. 2, p. 964-983, 2014.

LIMA, L. B. 2018. 164 p. **Análise da Qualidade da Cimentação de Poços de Petróleo Offshore Utilizando Técnicas de Mineração de Dados**. Dissertação (Mestrado em Engenharia Civil) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018.

LIN, J.; WILLIAMSON, S.; BORNE, K.; DEBARR, D. Pattern recognition in time series. **Advances in Machine Learning and Data Mining for Astronomy**, v. 1, p. 617-645, 2012.

LOBO, A.; ARA, F.; BARÓ, F.; CAMINO, C. Geospatial analysis for conservation: applications with open-source software in the Natural Parks of Barcelona. **Applied Geomatics**, v. 4, n. 2, p. 113-122, 2012.

LOCH, C. **A interpretação de imagens aéreas: noções básicas e algumas aplicações nos campos profissionais**. Florianópolis: Ed. UFSC, 1993.

LONGLEY, P. A.; GOODCHILD, M. F.; MAGUIRE, D. J.; RHIND, D. W. **Sistemas e ciência da informação geográfica**. 3º Ed. Bookman Editora. Porto Alegre. 2013, 540 p.

LORENZETT, C. D. C.; TELÖCKEN, A. V. Estudo Comparativo entre os algoritmos de Mineração de Dados Random Forests e J48 na tomada de Decisão. II Simpósio de Pesquisa e Desenvolvimento em Computação, v. 2, n. 1, p. 1-10, 2016.

MARSLAND, S.; **Machine Learning: An Algorithmic Perspective**. 1. ed. Chapman & Hall/CRC Press, 2008. 457 p.

MENESES, P. R.; SANO E. E. Classificação pixel a pixel de imagens. In: Meneses, P. R., Almeida, T. (Org.), **Introdução ao processamento de imagens de sensoriamento remoto**. 1º ed. Universidade de Brasília, Brasília, p. 191-208. 2012.

MERCIER, G; LENNON, M. Support vector machines for hyperspectral image classification with spectral based kernels. In: IEEE INTERNATIONAL GEOSCIENCE & REMOTE SENSING SYMPOSIUM, 23., 2003, Toulouse, France. **Proceedings...** Toulouse: GRSS, 2003. p. 288-290.

MITCHELL, T. **Machine Learning**. New York, McGraw Hill, 1997. 414 p.

MONARD, M. C.; BARANAUSKA, J. A. Indução de regras e árvores de decisão. In: REZENDE, S. O. (Org.). **Sistemas inteligentes-fundamentos e aplicações**. Barueri: Manole, 2003. p. 115-139.

MONTAÑO, R. A. N. R. **Aplicação de técnicas de aprendizado de máquina na mensuração florestal**. 2016. 102 p. Tese (Doutorado em Informática) - Universidade Federal do Paraná, Curitiba, Paraná, 2016.

MORAES, E. C. **Fundamentos do Sensoriamento Remoto**. 2002. Disponível em: <http://mtcm12.sid.inpe.br/col/sid.inpe.br/sergio/2005/06.14.12.18/doc/CAP1_ECMoraes.pdf>. Acesso em: 10 de jan. 2019.

MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto e metodologias de aplicação**. 3º ed. Viçosa: Ed. UFV, 2005. 320 p.

MOREIRA, M. A. **Fundamentos do Sensoriamento Remoto e Metodologias de aplicação**. 4º ed. São José dos Campos, São Paulo. 2011. 422 p.

MOREIRA, A. A.; NERY, C. V. M.; FERNANDES, F. H. S. Aplicação do algoritmo Support Vector Machine na análise espaço-temporal do uso e ocupação do solo na bacia do Rio Vieira. **Caminhos de Geografia**, v. 15, n. 50, 2014.

MOUNTRAKIS, G.; IM, J.; OGOLE, C. Support vector machines in remote sensing: A review. **Journal of Photogrammetry and Remote Sensing**, v. 66, p. 247-259, 2011.

MOURA, F. G. de. 2018. 28 p. **A dinâmica do uso e ocupação da terra no Lago Popóo (Bolívia) entre 1985 e 2017 utilizando classificação voltada ao objeto em dados landsat**. Monografia (Especialização em Geoprocessamento) – Universidade de Brasília, Brasília, DF, 2018.

MUELLER, J. P.; MASSARON, L. **Aprendizado de máquina para leigo**. Traduzido por João Tortello – Rio de Janeiro: Atlas Books, 2019. 482 p.

MÜLLER, H.; RUFIN, P.; GRIFFITHS, P.; BARROS SIQUEIRA, A. J.; HOSTERT, P. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. **Remote Sensing of Environment**, v. 156, p. 490-499, 2015.

NADER, R. S.; FAGGIN, C. A. M. **Geoprocessamento aplicado ao Urbanismo: introdução teórica e prática**. Apostila. São Paulo: Universidade de São Paulo, Faculdade de Arquitetura e Urbanismo, 2011.

NASCIMENTO, R. F. F.; ALCÂNTARA, E. H.; KAMPEL, M. STECH, J. L.; NOVO, E. M. L. M.; FONSECA, L. M. G. 2009. O algoritmo Support Vector Machines (SVM): Avaliação da separação ótima de classes em imagens CCD-CBERS-2. XIV Simpósio Brasileiro de Sensoriamento Remoto. **Anais...** Natal. 2009. p. 2079-2086.

NASA. **National Aeronautics and Space Administration**. 2019. Disponível em: <<http://landsat.gsfc.nasa.gov/about/tm.html>>. Acesso em: 10 de jan. 2019.

- NETO, C. D. G. **Potencial de técnicas de mineração de dados para o mapeamento de áreas cafeeiras**. INPE, São José dos Campos, 2014.
- NIKNEJAD, M.; MIRZAEI, Z. V.; HEYDARI, M. Comparing different classifications of satellite imagery in forest mapping (Case study: Zagros forests in Iran). **International Research Journal of Applied and Basic Sciences**, v.8, n. 9, p. 1407–1415, 2014.
- NOVO, E. M. L. de M. **Sensoriamento remoto: princípios e aplicações**. 3ª ed. São Paulo: Ed Blucher, 2008. 363 p.
- OLIVEIRA, J. V. T. M. de. 2018. 158 p. **Potencial de utilização de atributos espaciais no processo de classificação digital de plantios de seringueira no Mato Grosso**. Dissertação (Mestrado em Agronomia) – Universidade Estadual de Mato Grosso do Sul, Mato Grosso do Sul, 2018.
- OLIVEIRA, J. P. de. 2017. 148 p. **Detecção de áreas desmatadas na porção sul do estado do Amazonas, utilizando técnicas de extração de características e redes neurais artificiais**. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal do Amazonas, Manaus, Amazonas, 2017.
- OTUKEI, J. R.; BLASCHKE, T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. **International Journal of Applied Earth Observation and Geoinformation**, Enschede, v.12, supl. 1, p. 27–31, 2010.
- PAL, M. Random forest classifier for remote sensing classification. **International Journal of Remote Sensing**, v. 26, n. 1, p. 217-222, 2005.
- PASSO, D. P. 2013. 122 p. **Análise da qualidade de classificadores para identificação de alvos urbanos em imagens de alta resolução espacial: uma aplicação com as imagens do satélite Worldview II**. Dissertação (Mestrado em Geociências) – Universidade de Brasília, Brasília, Distrito Federal, 2013.
- PEREIRA, G. H. A.; CENTENO, J. A. S. Utilização de Support Vector Machine para classificação multiclases de imagens Landsat TM+. XVI Simpósio Brasileiro de Sensoriamento Remoto. **Anais...**Foz do Iguaçu. 2013. p. 2323-2328.
- PEREIRA, A. J.; SANTOS, A. M. Análise multitemporal da supressão vegetal no município de Xinguará-Pará. **Revista Eletrônica de Gestão e Tecnologias Ambientais**, v. 4, n. 2, p. 260-277, 2016.
- PMSBDF. **Plano Municipal de Saneamento Básico de Dona Francisca**. 2011. 133 p. Disponível em: <
http://www.donafrancisca.rs.gov.br/_arquivos/27a90292653be0ab00fdb43962f2bd22.pdf>. Acesso em: 10 de jan. 2019.

RIBEIRO, R. M. P.; SOARES, V. P.; VIEIRA, C. A. O. Avaliação de métodos de classificação de imagens IKONOS para o mapeamento da cobertura terrestre. XII Simpósio Brasileiro de Sensoriamento Remoto. **Anais...** Goiânia. 2005. p. 4277-4283.

ROSA, R. **Introdução ao Geoprocessamento**. 2013. Disponível em: <http://professor.ufabc.edu.br/~flavia.feitosa/cursos/geo2016/AULA5-ELEMENTOSMAPA/Apostila_Geop_rrosa.pdf>. Acesso em: 10 de jan. 2019.

ROCHA, C. H. B. **Geoprocessamento: Tecnologia Transdisciplinar**. Juiz de Fora: Ed. do Autor, 2000. 220 p.

ROY, D. P.; WULDER, M. A.; LOVELAND, T. R.; WOODCOCK, C. E.; ALLEN, R. G.; ANDERSON, M. C.; SCAMBOS, T. A. Landsat-8: Science and product vision for terrestrial global change research. **Remote sensing of Environment**, v. 145, p. 154-172, 2014.

RUIZ, L. F. C.; TEN CATEN, A.; DALMOLIN, R. S. D. Árvore de decisão e a densidade mínima de amostras no mapeamento da cobertura da terra. **Ciência Rural**, v. 44, n. 6, p. 1001-1007, 2014.

SANTOS, A. B. **Combinação de múltiplas abordagens de classificação para interpertação de imagens hiperespectrais de sensoriamento remoto**. 2013. 103 p. Dissertação (Ciências da Computação) - Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, 2013.

SANTOS, P. A. DOS. 2014. 80 p. **Avaliação do uso de variáveis temporais na classificação da cobertura da terra**. Dissertação (Ecologia) - Universidade Federal de Lavras, Lavras, Minas Gerais, 2014.

SANTOS, G. D. dos.; OLIVEIRA, R. A. A. C. Classificação pixel a pixel aplicada à discriminação de corpos d'água em imagem landsat 8. **Revista Militar de Ciência e Tecnologia**, v. 33, n. 1, p. 5-8, 2016.

SARMIENTO, C. M.; RAMIREZ, G. M.; COLTRI, P. P.; SILVA, L. F. L.; NASSUR, O. A. C.; SOARES, J. F. Comparação de classificadores supervisionados na discriminação de áreas cafeeiras em Campos Gerais – Minas Gerais. **Coffee Science**, Lavras, v. 9, n. 4, p. 546- 557, 2014.

SHWARTZ-SHALEV, S.; BEN-DAVID, S. **Understanding Machine Learning: From theory to algorithms**. NY, USA: Cambridge University Press, 2014.

SILVA, D. A. da. **Sistemas Sensores Orbitais**. São José dos Campos: INPE – CTA, 1995. 69 p.

SILVA, F. C. **Análise ROC (Receiver Operating Characteristic)**. São José dos Campos, 2006.

SILVA, E. A. **Modelagem dinâmica para simulação no processo de arenização e cobertura florestal na Campanha Ocidental – RS**. 2015. 163 p.

Tese (Doutorado em Engenharia Florestal) – Universidade Federal de Santa Maria, Rio Grande do Sul, 2015.

SILVA, A. M. L.; FILHO, F. das C. I.; JUNIOR, J. M. de S.; MACHADO, V. P. Descoberta de conhecimento através de métodos de aprendizagem de máquina supervisionados aplicados ao SIGAA/UFPI. **Revista de Sistemas e Computação**, v. 7, n. 1, 2017.

SIVIERO, M, R. L. JÚNIOR, E. R. H. aplicação de aprendizado de máquina em dados mensurados numa seção do Rio Atibaia/SP. **Anais... 30º Congresso Nacional de Saneamento e Meio Ambiente e 30º Feira Nacional de Saneamento e Meio Ambiente**. p. 1-5, 2019.

SMOLA, A. J.; BARLETT, P.; SCHÖLKOPF, B.; SCHUURMANS, D. **Advances in Large Margin Classifiers**. Cambridge: MIT Press, 2000. 423 p.

SOLIMAN, O.; MAHMOUD, A. A classification system for remote sensing satellite images using support vector machine with non-linear kernel functions. In: **Informatics and Systems (INFOS)**, 2012 8th International Conference on. [S.l.: s.n.], 2012. p. 181-187.

SOUSA, B. F. S. 2009. 90 p. **Aprendizado de máquina na detecção do uso do solo no Bioma Caatinga via Sensoriamento Remoto**. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Federal do Ceará, Fortaleza, Ceará, 2009.

SOUSA, B. F. S.; TEIXEIRA, A. S.; SILVA, F. A. T. F.; ANDRADE, E. M.; BRAGA, A. P. S. Avaliação de classificadores baseados em aprendizado de máquina para a classificação do uso e cobertura da terra no Bioma Caatinga. **Revista Brasileira de Cartografia**, n. 62, 2010.

SOUZA, J. F.; KAI, P. M. **Classificação de folhas usando medidas invariantes**. Monografia - Curso de Ciência da Computação - Universidade Estadual de Mato Grosso do Sul, 113 p, 2014.

SOUZA, C. G. 2015. 162 p. **Uso de séries temporais para o mapeamento da cafeicultura**. Tese (Doutorado em Engenharia Florestal) – Universidade Federal de Lavras, Minas Gerais, 2015.

SOUZA, C. G.; CARVALHO, L.; AGUIAR, P.; ARANTES, T. B. Algoritmo de aprendizagem de máquina e variáveis de Sensoriamento Remoto para o mapeamento da cafeicultura. **Boletim de Ciências Geodésicas**, v. 22, n. 4, 2016.

STEINIGER, S.; HAY, G. J. Free and open source geographic information tools for landscape ecology. **Ecological Informatics**, v. 4, n. 4, p. 183-195, 2009.

STRECK, E. V.; KÄMPF, N.; DALMOLIN, R. S. D.; KLAMT, E.; NASCIMENTO, P. C.; SCHNEIDER, P.; GIASSON, E.; PINTO, L. F. S. **Solos do Rio Grande do Sul**. 2º ed. ver. e ampl. Porto Alegre: Emater/RS, 2008. p. 222.

SU, M.Y. Using clustering to improve the KNN-based classifiers for online anomaly network traffic identification. **Journal of Network and Computer Applications**, v. 34, n. 2, p. 722-730, 2011.

TAKAKURA, A. M.; PEREIRA, R. D.; ASSIS DA SILVA, F., PAZOTI, M. A.; DE ALMEIDA, L. I.; MOLINA, H. S. Uso do aprendizado de máquina no diagnóstico médico de patologias. **Colloquium Exactarum**, v. 10, n.1, p. 78-90, 2018.

TATSUMI, K.; YAMASHIKI, Y.; TORRES, M. A. C.; TAIPE, C. L. R. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. **Computers and Eletronics in Agriculture**, v. 115, p. 171-179, 2015.

TIAN, S.; ZHANG, X.; TIAN, J.; SUN, Q. Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China. **Remote Sensing**, v. 8, n. 11, p. 954, 2016.

TOPALOĞLU, R. H.; SERTEL, E.; MUSAOĞLU, N. Assessment of classification accuracies of sentinel-2 and landsat-8 data for land cover/use mapping. XXIII International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, ISPRS. **Anais...Prague, Czech Republic**, v. XLI-B8, 2016. p. 1055-1059.

VENDRUSCOLO, C. E. 2010. 166 p. **As relações do espaço urbano do município de Dona Francisca, RS com a região da Quarta Colônia**. Dissertação (Mestrado em Geografia) - Universidade Federal de Santa Maria, Santa Maria, Rio Grande do Sul, 2010.

VIDHYA, M. Efficient classification of portscan attacks using support vector machine. In: **Green High Performance Computing (ICGHPC)**, 2013 IEEE International Conference on. [S.l.: s.n.], 2013. p. 1-5.

VIEIRA, C. A. O. **Accuracy of remotaly sensing classification of agricultural crops: a comparative study**. 2000. 327p. Thesis (Doctor of Philosophy). University of Nottingham, 2000, p. 128-175.

WANG, H.; MA, C.; ZHOU, L. A Brief Review of Machine Learning and its Application. In: **Information Engineering and Computer Science, 2009. ICIECS**. International Conference on. IEEE, 2009. 1-4 p.

WANG, X.; GAO, X.; ZHANG, Y.; FEI, X.; CHEN, Z.; WANG, J.; ZHAO, H. Land-Cover classification of coastal wetlands using the RF algorithm for Worldview-2 and Landsat 8 images. **Remote Sensing**, v. 11, n. 16, p. 1927, 2019.

WIELAND, M.; PITTORE, M. Performance evaluation of machine learning algorithms for urban pattern recognition from multi-spectral satellite images. **Remote Sensing**, Nova Zelândia, v. 6, n. 4, 2912–2939, 2014.

XU, Y.; ZHU, Q.; FAN, Z.; QIU, M.; CHEN, Y.; LIU, H. Coarse to fine K nearest neighbor classifier. **Pattern Recognition Letters**, North-Holland, v. 34, n. 9, p. 980-986, 2013.

ZANATA, J. M. 2014. 122 p. **Mudanças no uso e cobertura da terra na bacia hidrográfica do Ribeirão Bonito, municípios de Avaré e Itatinga-SP.** Dissertação (Mestrado em Geografia) – Universidade Estadual Paulista, Presidente Prudente, São Paulo, 2014.

ZHAI, L.; SUN, J.; SANG, H.; YANG, G.; JIA, Y. Large area land cover classification with Landsat ETM+ images based on decision tree. **Anais...** International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, v. XXXIX, 2012.

ZHU, Z.; WOODCOCK, C. E.; ROGAN, J.; KELLNDORFER, J. Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. **Remote Sensing of Environment**, v. 117, p. 72-82, 2012.