



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
Programa de Pós-Graduação em Engenharia Ambiental  
Câmpus Apucarana e Londrina



PRISCILA GABRIELA DA SILVA DUARTE

**MÉTODOS DE *MACHINE LEARNING* APLICADOS À  
CLASSIFICAÇÃO DO USO E OCUPAÇÃO DO SOLO NA  
MICROBACIA DO LAGO IGAPÓ NA CIDADE DE LONDRINA/PR**

**LONDRINA**

**2022**

PRISCILA GABRIELA DA SILVA DUARTE

**MÉTODOS DE *MACHINE LEARNING* APLICADOS À  
CLASSIFICAÇÃO DO USO E OCUPAÇÃO DO SOLO NA  
MICROBACIA DO LAGO IGAPÓ NA CIDADE DE LONDRINA/PR**

Machine learning methods applied to the classification of soil use and occupation in the microbain of Igapó Lake in the city of Londrina/PR.

Dissertação apresentada como requisito parcial à obtenção do grau de mestre em Engenharia Ambiental, ao Programa de Pós-graduação em Engenharia Ambiental, da Universidade Tecnológica Federal do Paraná. Área de concentração: Processos Atmosféricos.

Orientador: Prof. Dr. Jorge Alberto Martins.  
Coorientador(a): Prof(a). Dr(a). Ligia Flávia Antunes Batista.

**LONDRINA**

**2022**



Esta licença permite compartilhamento do trabalho, mesmo para fins comerciais, sem a possibilidade de alterá-lo, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



Ministério da Educação  
Universidade Tecnológica Federal do Paraná  
Campus Londrina



PRISCILA GABRIELA DA SILVA DUARTE

**MÉTODOS DE MACHINE LEARNING APLICADOS À CLASSIFICAÇÃO DO USO E OCUPAÇÃO DO SOLO  
NA MICROBACIA DO LAGO IGAPÓ NA CIDADE DE LONDRINA/PR**

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre em Engenharia Ambiental, da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Engenharia Ambiental.

Data de aprovação: 05 de Agosto de 2022

Dr. Jorge Alberto Martins, Doutorado - Universidade Tecnológica Federal do Paraná

Dra. Adriana Castreghini De Freitas Pereira, Doutorado - Universidade Estadual de Londrina (UEL)

Dr. Mauricio Moreira Dos Santos, Doutorado - Universidade Tecnológica Federal do Paraná

Thiago Gentil Ramires, - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 05/08/2022.

*Dedico a Deus, a meus pais Márcia e Adilson,  
meus irmãos Douglas e Maria Tereza ao meu primo  
Eder e minha tia Patricia e todos familiares que me  
ajudaram nessa conquista.*

## AGRADECIMENTOS

Agradeço primeiramente a Deus pelo dom a vida e todo seu amor por nós. Agradeço em especial minha mãe Marcia Helena da Silva, uma mulher incrível que nunca me deixou desistir e sempre me incentivou a estudar e procurar o melhor caminho, agradeço também meu pai Adilson Rafael Duarte, meus irmãos Douglas Silva Duarte e Maria Tereza Silva Duarte, em memória a minha avó Tereza Rafael Duarte em memória, e todos que participaram e contribuíram na minha jornada para a realização a conquista do meu título de Mestre(a) em Engenharia Ambiental. Agradeço por todos os conselhos e apoio no decorrer dessa jornada. Agradeço a Deus por ter me proporcionado a oportunidade de cursar a pós-graduação em uma excelente universidade Federal como a Universidade Tecnológica Federal do Paraná campus Londrina. Agradeço a Prof. Dr<sup>o</sup> Jorge Alberto Martins pela orientação durante esses dois anos e pela oportunidade de fazer parte do grupo de pesquisa, neste período e por ele abraçar a ideia para eu concretizar a minha dissertação de mestrado. Agradeço a Prof. Dr<sup>a</sup> Ligia Flavia Antunes Batista, pela sua coorientação e imensa colaboração para eu desenvolver a dissertação inclusive alcançar meus resultados, seus ensinamentos me deram norte para conduzir a pesquisa de forma produtiva, agradeço ambos professores por todo ensinamento o qual me proporcionaram. Agradeço aos meus colegas por toda ajuda que me ofereceram durante o período de mestrado, em especial a minha amiga de pesquisa Pamela Venturin. Agradeço a oportunidade de participar do protejo *Rain Solutins* e Fundação Araucária por ter investido em minha pesquisa. O segredo da vida não está na chegada na linha final e ganhar o lugar mais alto do pódio, e sim, tudo aquilo que levamos para si, mediante a tudo que absorvemos das pessoas e, o que emitimos para elas da nossa essência durante esse percurso.

DUARTE, Priscila Gabriela Silva. **Métodos de *Machine Learning* Aplicados à Classificação do Uso e Ocupação do Solo Na Microbacia Do Lago Igapó Na Cidade De Londrina/Pr.** 2022. 77 p. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Engenharia Ambiental, Universidade Tecnológica Federal do Paraná (UTFPR), Londrina, 2022.

## RESUMO

O crescimento das áreas urbanas é um fenômeno mundial e seus problemas têm gerado imensos desafios para a sociedade. Um desses desafios está relacionado à drenagem das águas de chuva em ambiente urbano. Com a urbanização, a cobertura original do solo é substituída por elementos que impedem ou dificultam a infiltração das águas da chuva. As soluções baseadas na natureza têm se mostrado uma alternativa viável para o enfrentamento dos problemas associados à drenagem urbana. Entretanto, o sucesso das intervenções é altamente dependente de uma compreensão dos elementos que predominam na superfície urbana. Nesse sentido, o principal objetivo deste trabalho foi a aplicação de algoritmos de *machine learning* em imagens de satélite de alta resolução para classificação do uso e ocupação do solo na área de drenagem do sistema de lagos Igapó, um conjunto de lagos urbanos da cidade de Londrina-PR. Imagens da constelação de satélites Pléiades, com elevado nível de detalhamento espacial, foram usadas para a classificação. Foram estabelecidas 12 classes temáticas específicas para classificação do uso do solo. A classificação supervisionada foi aplicada e os seguintes algoritmos de aprendizado de máquina foram avaliados: *Decision Tree* (DT), *Randon Forest* (RT), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN) e *Normal Bayes*. DT foi o classificador que apresentou o melhor desempenho, tanto para a classificação global quanto para as classes individuais. Os valores para os índices Kappa, Precisão, Recall e F1-Score ficaram entre 90 % e 100 % para o classificador DT. As áreas verdes, representadas por Árvores/arbustos e Grama, juntamente com telhados, representam as classes majoritárias, com 24% e 23% da cobertura do solo da área de estudo, respectivamente. Asfalto e calçada, com 19% da área, e solo exposto, com 14%, também foram classes com participações expressivas. A fração remanescente é constituída por água, inclusive superfície de piscina, e áreas sombreadas. Com o refinamento das classes, a resolução espacial e a qualidade do mapeamento realizado, os resultados apresentados neste trabalho podem ser uma ferramenta de grande utilidade para a elaboração de projetos de intervenção que promovam o uso de soluções baseadas na natureza para a solução dos problemas relacionados à drenagem urbana.

Palavras-chave: Soluções baseada na Natureza; Aprendizagem de Máquina, Árvore de Decisão; Planejamento Urbano; Sensoriamento Remoto.

DUARTE, Priscila Gabriela Silva. **Machine learning methods applied to the classification of soil use and occupation in the microbain of Igapó Lake in the city of Londrina/PR.** 2022. 77 p. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Engenharia Ambiental, Universidade Tecnológica Federal do Paraná (UTFPR), Londrina, 2022.

### ABSTRACT

The growth of urban areas is a worldwide phenomenon and its problems have generated immense challenges for society. One of these challenges is related to the drainage of rainwater in an urban environment. With urbanization, the original land cover is replaced by elements that prevent or hinder the infiltration of rainwater. Nature-based solutions have proved to be a viable alternative for dealing with problems associated with urban drainage. However, the success of interventions is highly dependent on an understanding of the elements that predominate on the urban surface. In this sense, the main objective of this work was the application of machine learning algorithms in high resolution satellite images to classify land use and occupation in the drainage area of the Igapó lake system, a set of urban lakes in the city of Londrina-PR. Images of the Pléiades satellite constellation, with a high level of spatial resolution, were used for the classification. Twelve specific thematic classes were established for classifying land cover. Supervised classification was applied, and the following Machine Learning algorithms were evaluated: Decision Tree (DT), Random Forest (RT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Normal Bayes. DT was the classifier that presented the best performance, both for the global classification and for the individual classes. The values for the Kappa, Precision, Recall, and F1-Score indices were between 90% and 100% for the DT classifier. The green areas, represented by Trees/bushes and Grass, together with roofs, represent the majority classes, with 24% and 23% of the land cover of the study area, respectively. Asphalt and pavement, with 19% of the area, and exposed soil, with 14%, were also classes with expressive participation. The remaining fraction consists of water, including pool surface, and shaded areas. With the refinement of the classes, the spatial resolution, and the quality of the developed mapping, the results presented in this work can be a very useful tool for the elaboration of intervention projects that promote the use of nature-based solutions to solve urban drainage problems.

Keywords: Nature-based solutions; Machine Learning, Decision tree; Urban planning; Remote sensing.

## LISTA DE ILUSTRAÇÕES

<b>Figura 1 - Interação Radiação Eletromagnética.....</b>	<b>18</b>
<b>Figura 2 - Representação do SVM. ....</b>	<b>24</b>
<b>Figura 3 - Estrutura Clássica de uma árvore de Decisão .....</b>	<b>26</b>
<b>Figura 4 - Mapa de localização do município de Londrina – PR.....</b>	<b>32</b>
<b>Figura 5 - Mapa de Solos de Solos da Microbacia do Ribeirão Cambé-PR.....</b>	<b>33</b>
<b>Figura 6 - Mapa de Localização Bacia Ribeirão Cambé. ....</b>	<b>35</b>
<b>Figura 7 - Mapa de Localização Lago Igapó na sub-bacia do Rio Cambé no município de Londrina-PR. ....</b>	<b>36</b>
<b>Figura 8 - Fluxograma de delimitação de bacia através de ponto exultório.....</b>	<b>38</b>
<b>Figura 9 - Fluxograma das etapas do processamento digital de imagem (Pléiades). ....</b>	<b>39</b>
<b>Figura 10 - Imagem de alta resolução com área delimitada da área de estudo em Londrina-PR, no ano de 2020.....</b>	<b>40</b>
<b>Figura 11 - Amostras de Treinamento (treino e validação) para as Classes estabelecidas. ....</b>	<b>45</b>
<b>Figura 12 - Uso do Stretview para confirmação da classificação.....</b>	<b>49</b>
<b>Figura 13 - Segmentação de imagem no TerraView. ....</b>	<b>50</b>
<b>Figura 14 - Amostra de cada classificador. ....</b>	<b>51</b>
<b>Figura 15 - Métricas de acurácia. ....</b>	<b>52</b>
<b>Figura 16- Acurácia Global. ....</b>	<b>56</b>
<b>Figura 17 - Mapa de classificação de uso e ocupação do solo modelo Decision Tree. ....</b>	<b>57</b>
<b>Figura 18 - Porcentagem das Subclasses de Uso e Ocupação do Solo .....</b>	<b>13</b>
<b>Figura 19 - Áreas com potencial de infiltração.....</b>	<b>14</b>



## LISTA DE TABELAS

<b>Tabela 1: Especificações técnicas da imagem de satélite da área de interesse.....</b>	<b>38</b>
<b>Tabela 2: Valores de teste para segmentação de imagem no TerraView. ....</b>	<b>41</b>
<b>Tabela 3: Campos Computados pelo Algoritmo ZonalStatistic (OTB).....</b>	<b>42</b>
<b>Tabela 4: Grau de exatidão da classificação de acordo com o valor do Índice Kappa. ...</b>	<b>47</b>

## LISTA DE QUADROS

<b>Quadro 1: Classes de elementos para determinação de uso e ocupação do solo .....</b>	<b>42</b>
<b>Quadro 2: Matriz de Confusão Classificação DT.....</b>	<b>46</b>

## LISTA DE SIGLAS E ABREVIATURAS

AGUASPARANÁ	Instituto das Águas do Paraná
AM	Aprendizado de Máquina
AR	Análises de Regressão
ARO	Avaliação de Risco Operacional
B	<i>Blue</i>
Bits	<i>Binary digits</i>
CBERS	<i>China-Brazil Earth Resources Satellite</i>
Cfa	Clima Temperado Húmido com verão quente
CE	Condutividade Elétrica
cm	Centímetro
CMOS	<i>Complementary Metal Oxide Semiconductor</i>
DECEA	Departamento de Controle do Espaço Aéreo
DQO	Demanda Química de Oxigênio
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
<i>et al.</i>	Outros autores
FA	Floresta Aleatória
G	<i>Green</i>
g	Gramas
GLONASS	<i>Global Navigational Satellite System</i>
GNSS	<i>Global Navigation Satellite System</i>
GPS	<i>Global Positioning System</i>
IA	Inteligência Artificial
IAP	Instituto Ambiental do Paraná
IAPAR	Instituto Agrônômico do Paraná
IAT	Instituto de Água e Terra
IBGE	Instituto Brasileiro de Geografia e Estatística
INPE	Instituto Nacional de Pesquisas Espaciais
IP67	A prova de poeira totalmente vedado
IP68	Protegido contra imersão contínua em água
ITCG	Instituto de Terras, Cartografia e Geologia do Paraná
kg	Quilograma

km	Quilômetro
km/h	Quilômetro por hora
km <sup>2</sup>	Quilômetro quadrado
LANDSAT	<i>Land Remote Sensing Satellite</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>13</b>
<b>2 OBJETIVO</b> .....	<b>16</b>
<b>2.1 Objetivo geral</b> .....	16
<b>2.2 Objetivos específicos</b> .....	16
<b>3. REVISÃO DA LITERATURA</b> .....	<b>17</b>
<b>3.1 Sensoriamento Remoto</b> .....	17
<b>3.2 Geoprocessamento</b> .....	19
<b>3.3 Classificação de imagens digitais</b> .....	20
3.3.1 Classificação baseada a pixel .....	21
3.3.2 Classificação baseada em objetos .....	22
<b>3.4 Aprendizagem de Máquina (<i>Machine Learning</i>)</b> .....	22
3.4.1 Modelos de aprendizagem de <i>Machine Learning</i> .....	23
3.4.2 Máquinas de Vetores de Suporte ( <i>Support Vector Machine – SVM</i> ) .....	24
3.4.3 Árvores de decisão ( <i>Decision Tree – DT</i> ).....	25
3.4.4 Florestas aleatórias ( <i>Random Forest – RF</i> ) .....	26
3.4.5 K-Vizinhos mais próximos ( <i>K-Nearest Neighbors - KNN</i> ) .....	26
3.4.6 <i>Naive Bayes – NB</i> .....	27
<b>3.5 Uso e Cobertura do Solo</b> .....	27
3.5.1 Caracterização do Tema .....	27
3.5.2 Mapeamento da cobertura do solo .....	29
<b>3.6 Soluções Baseadas na Natureza (SbN)</b> .....	30
<b>4 METODOLOGIA</b> .....	<b>32</b>
<b>4.1 Caracterização da área de estudo</b> .....	32
4.1.1 Área de estudo .....	32
4.1.2 Clima .....	33
4.1.3 Solos .....	33

4.1.4 Vegetação .....	34
4.1.5 Hidrografia .....	34
4.1.6 Lago Igapó.....	36
<b>4.2 Material .....</b>	<b>37</b>
<b>4.3 Delimitação da sub bacia do Ribeirão Cambé .....</b>	<b>37</b>
<b>4.4 Aquisição de imagem de alta resolução .....</b>	<b>38</b>
<b>4.5 Classificação da imagem .....</b>	<b>41</b>
4.5.1 Segmentação de da imagem .....	41
4.5.2 Extração de atributos .....	41
4.5.3 Definição das Classes de uso e ocupação do solo e seleção de amostras de treinamento e validação.....	42
4.5.4 Treinamento e validação.....	45
4.5.5 Classificação.....	46
4.5.6 Avaliação dos modelos.....	46
<b>4.6 Confirmação da exatidão do classificador <i>Streetview</i>.....</b>	<b>48</b>
<b>5 RESULTADOS E DISCUSSÃO.....</b>	<b>50</b>
<b>5.1 Desempenho dos modelos.....</b>	<b>51</b>
<b>5.2 Mapeamento da cobertura do solo para a BH .....</b>	<b>54</b>
<b>5.3 Potenciais Aplicações do Estudo .....</b>	<b>14</b>
<b>6 CONCLUSÃO.....</b>	<b>16</b>
<b>REFERÊNCIAS.....</b>	<b>17</b>
<b>APÊNDICES.....</b>	<b>27</b>

## 1 INTRODUÇÃO

O desenvolvimento acelerado da urbanização, muitas vezes sem planejamento ou com planejamento inadequado, acarreta degradação dos recursos naturais, como vegetação, cursos d'água e solos. Isso ocorre devido à busca de espaço para a construção de obras de engenharia, sendo inseridas, em muitos casos, em locais impróprios decorrentes da falta de planejamento. Essa inadequação ocasiona e pode acarretar severos problemas futuros, como aqueles presenciados atualmente, como alagamento, enchentes, assoreamento de lagos e rios, ilhas de calor, entre outros.

A área de estudo encontra-se localizada em um perímetro urbano no município de Londrina-PR, a qual passou por intensa antropização nas últimas décadas, acarretando problemas de planejamento urbano devido à expansão desordenada. A área de interesse sendo a microbacia do Ribeirão Cambé conta com um complexo de 4 lagos dentro de sua área de drenagem, conhecido como Lago Igapó I, II, III e IV, sendo eles os alvos principais para o presente estudo.

Larroca et al. (2017), em seu estudo onde avalia os impactos ambientais decorrentes das ações humanas, identificaram que os principais efeitos com relação ao meio geofísico no entorno do Lago Igapó foram atribuídos principalmente em relação à impermeabilização do solo, decorrentes das vias e da construção civil. Um dos principais impactos negativos tem sido o carregamento de resíduos para o curso d'água, principalmente pelo bairro Gleba Palhano, que está em um processo de consolidação. Outro fato preocupante é o transporte de resíduos, que acarreta o assoreamento do lago.

Nesse sentido, a tecnologia pode ser uma grande aliada do planejamento urbano, pois auxilia em tomadas de decisões e colabora com a otimização de tempo. Ferramentas como, por exemplo, imagem de satélite de alta resolução, técnicas de *Machine Learning*, entre outros, fazem parte dessa contribuição tecnológica. Com isso, o presente trabalho utilizou de tais ferramentas para a análise de uso e ocupação do solo na Sub bacia do Ribeirão Cambé, por classificação supervisionada orientada por objeto, utilizando algoritmos de ML para a classificação.

Nandaia (2020), relata que os avanços científicos registados no domínio da classificação de imagens baseada em objeto, desde o início do século, confirmaram esta técnica como uma das mais promissoras no que diz respeito à extração de informação de uso e ocupação do solo.

Como a antropização é um processo praticamente irreversível, uma forma de entender as condições atuais de certo local, e colaborar para minimizar os processos de degradação dos

elementos naturais, é analisar as condições de uso e ocupação do solo. O entendimento dessas condições permite identificar áreas com grande degradação e com problemas associados, os quais são objetos de estudo recorrentes do planejamento urbano.

O uso e ocupação do solo tem influência direta sobre diversos processos ambientais, dentre os quais citam-se os efeitos deletérios sobre a qualidade e a disponibilidade dos recursos hídricos de uma bacia hidrográfica. Assim, o mapeamento do uso e cobertura do solo surge como uma ferramenta de análise ambiental com o intuito de visualizar as perdas de biodiversidade, vegetação, recursos hídricos, alterações climáticas e incidência de doenças, além do impacto social causado pela urbanização (SCHNEIDER et al., 2016).

Para alcançar o mapeamento e avaliação do uso e ocupação do solo, o uso de ferramentas computacionais como *Machine Learning* e geotecnologias, conjuntamente aplicados na manipulação de dados de sensoriamento remoto, gera produtos adequados para um bom planejamento de ações de manejo que contribuem para a solução de problemas e promovem melhorias na qualidade de vida e dos recursos naturais.

Para Souza (2017), as tecnologias e metodologias de processamento de imagens de sensoriamento remoto apresentam um potencial indispensável nas análises ambientais, pois fornecem informações sobre a cobertura da terra e sua dinâmica, e seus produtos conseguem monitorar as mudanças na paisagem, em dimensões temporais e espaciais. Com isso, ressalta-se que mapeamentos de uso e ocupação do solo são de grande valor para o planejamento urbano, visando a mitigação dos problemas gerados pelo crescimento urbano, muitas vezes sem planejamento adequado.

Métodos de *Machine Learning* podem ajudar a processar uma ampla variedade de conjuntos de dados de sensoriamento remoto. Através desses métodos é possível processar diferentes tipos de dados de sensoriamento remoto. Suas aplicações têm demonstrado alta capacidade ao combinar dados de sensoriamento remoto e métodos de aprendizagem de máquina para mapear diferentes recursos naturais (SHIRMARD, et. al, 2022).

Com o crescimento e desenvolvimento do município de Londrina, foram surgindo problemas de drenagem de águas pluviais, transporte de sedimento para dentro do lago Igapó, acarretando seu assoreamento. Para minimizar esses impactos, a aplicação de soluções baseadas na natureza SbN é um método eficaz para implementar elementos naturais para a mitigação de problemas ambientais no meio urbano. Entretanto, um bom mapeamento das condições de uso e ocupação do solo é o primeiro passo para identificar quais intervenções SbN são as mais adequadas para uma região de interesse.



Soluções baseadas na Natureza (SbN), ou em inglês *Nature-based Solutions*, é uma expressão que surge como um conceito polifônico e detentor das demais estratégias consolidadas no âmbito de serviços ecossistêmicos e de capital natural. Ou seja, a ideia em torno das SbN funciona como um conceito guarda-chuva que busca expressar todas as soluções que, de alguma forma, se inspiraram, copiam ou tomam como base processos naturais para gerar algum benefício para o ambiente e para a sociedade humana. (FRAGA e SAYAGO, 2020).

O presente estudo avaliou as condições de uso e ocupação do solo da parte alta da microbacia do Ribeirão Cambé, mais especificamente a área de drenagem do sistema de lagos Igapó, da área urbana do município de Londrina-PR. Com os resultados, espera-se contribuir para um aproveitamento mais eficiente das SbN, para que se tornem parte do planejamento urbano futuro da área de estudo.

## 2 OBJETIVO

### 2.1 Objetivo geral

O objetivo do presente estudo é avaliar o uso e ocupação do solo na microbacia do Ribeirão Cambé, onde se localiza a área de drenagem do Lago Igapó, na área urbana do município de Londrina – PR, a partir de imagens de satélite com elevado nível de detalhamento espacial, utilizando técnicas de *Machine Learning* através de classificação supervisionada orientada por objeto.

### 2.2 Objetivos específicos

- Identificar e adquirir uma base de dados de imagem de satélite de altíssima resolução para a área de drenagem do lago Igapó;
- Definir classes de uso e ocupação do solo adequadas à área do lago e às necessidades do planejamento urbano na aplicação de SbN;
- Aplicar técnicas de *Machine Learning* para classificação e quantificar o uso e ocupação do solo por meio de classificação supervisionada das imagens de satélite adquiridas;
- Contribuir, através de um mapeamento do uso e ocupação do solo, com a identificação de áreas com potencial de infiltração de águas pluviais, para um melhor planejamento urbano utilizando SbN.

### 3. REVISÃO DA LITERATURA

#### 3.1 Sensoriamento Remoto

O desenvolvimento inicial do sensoriamento remoto é cientificamente ligado ao desenvolvimento da fotografia e à pesquisa espacial. As fotografias aéreas foram o primeiro produto de sensoriamento remoto a ser utilizado. O termo sensoriamento remoto (SR) apareceu pela primeira vez na literatura científica em 1960, e desde então esse termo tem abrigado tecnologia e conhecimentos extremamente complexos derivados de diferentes campos que vão desde a física até a cartografia (NOVO e PONZONI, 2001).

Segundo Santos (2013), o sensoriamento remoto é uma tecnologia que utiliza sensores para extrair informações sobre objetos ou fenômenos, a partir da radiação eletromagnética (REM), sem que ocorra contato direto entre eles, ou seja, sem que o homem seja parte essencial do processo de aquisição dos dados.

O sensoriamento remoto consiste na utilização conjunta de sensores, equipamentos para o processamento e transmissão de dados da plataforma (aérea ou espacial), com o objetivo de estudar o ambiente terrestre através de registro e análises das interações entre a radiação eletromagnética e substâncias componentes do planeta em suas variadas manifestações (NOVO, 1989).

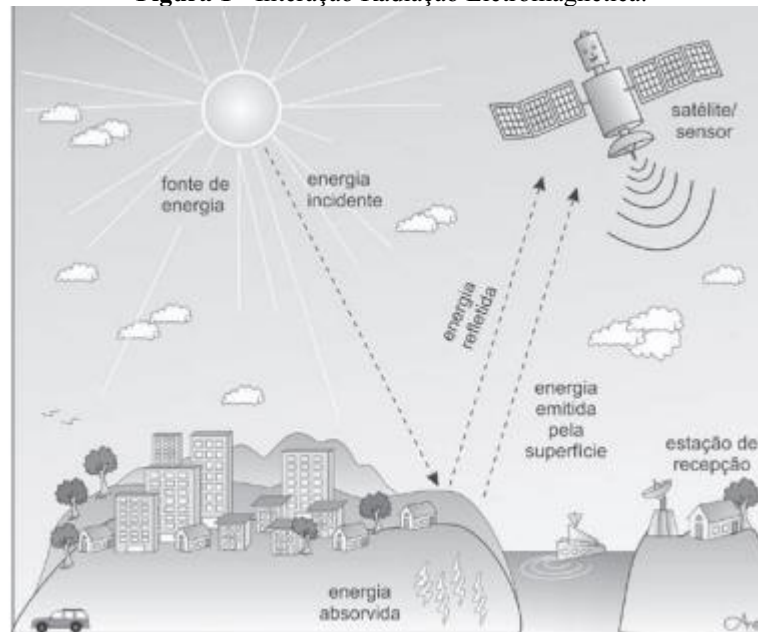
Para Lillesand e Kiefer (1994), o sensoriamento remoto é a ciência que permite obter informação sobre um objeto (alvo), área ou fenômeno através da análise de dados adquiridos por um dispositivo (sensor) que não está em contato direto com o objeto, área ou fenômeno sob investigação. O sensoriamento remoto realiza observações do planeta através de sensores acima do solo. Esses sensores podem ser câmeras, que detectam não somente a luz visível, mas também a radiação de outros comprimentos de onda, como o infravermelho, por exemplo (SOUZA, 2010).

Sensores que obtém imagens que não sejam pela detecção da radiação eletromagnética (REM) também devem ser classificados como relacionados ao sensoriamento remoto. A confusão mais comum é feita com os sensores geofísicos aerotransportados, tais como os magnetômetros, que geram uma imagem a partir de medições do campo magnético da superfície terrestre. Portanto, neste caso não há qualquer relação com a energia eletromagnética (MENESES e ALMEIDA, 2012).

A partir da definição de sensoriamento remoto, fica evidente que é necessário conhecer o que é a REM para que se possa entender o processo de aquisição e de análise das imagens

(MENESES e ALMEIDA, 2012). A REM começa com o entendimento da dualidade do comportamento da sua natureza: onda e energia. A REM se propaga pelo espaço vazio por meio de um campo eletromagnético variável (Figura 1). A energia produzida pelo Sol, por exemplo, chega à Terra por esse processo de transmissão. A luz solar é a forma mais conhecida de radiação

**Figura 1** - Interação Radiação Eletromagnética.



Fonte: Quartaroli, et. al (2014).

A REM é o elemento de ligação entre todos os demais. Para o caso de aplicação das técnicas de sensoriamento remoto no estudo dos recursos naturais, tanto o Sol quanto a Terra podem ser a fonte de REM. Embora a terra não emita luz visível, ela é fonte para os sensores passivos de micro-ondas e termais (NOVO e PONZONI, 2001).

O sensor é o instrumento capaz de coletar e registrar a REM refletida ou emitida pelo objeto, que também é denominado alvo, e que representa o elemento do qual se pretende extrair informação (MENESES e ALMEIDA, 2012). Ele também é capaz de responder a faixas específicas do espectro eletromagnético, registrá-la e gerar um produto numa forma adequada para a interpretação do usuário (STEFFEN et al., 1981). Os sensores remotos podem ser classificados em função da fonte de energia. Quando os sensores não possuem fonte própria de radiação são denominados de sensores passivos (e.g., sensores que medem a radiação solar refletida pela superfície da Terra ou as ondas de infravermelho emitidas por ela). Por outro lado, quando os sensores possuem sua própria fonte de REM são denominados sensores ativos. Radares, por exemplo, são dispositivos que medem a radiação que eles próprios emitem, mas que é refletida pelo alvo (INPE, 2006).

Os sensores podem ainda ser classificados em função do tipo de produto. Sensores não-imageadores são aqueles que não geram imagem da superfície sensoriada (e.g., radiômetros e espectro-radiômetros). Sensores não-imageadores são essenciais para aquisição de informações precisas sobre o comportamento espectral dos objetos. Por outro lado, sensores imageadores produzem uma imagem, fornecendo informações sobre a variação espacial da resposta espectral da superfície observada (INPE, 2006).

### **3.2 Geoprocessamento**

O geoprocessamento, segundo Câmara et al. (2001), consiste no conhecimento que utiliza técnicas matemáticas e computacionais para o tratamento da informação geográfica. Isso envolve técnicas e métodos que vão desde a coleta, passando pelo tratamento e armazenamento, até o processamento dos dados, com a finalidade de produzir novos conhecimentos sobre o assunto abordado. Esta tecnologia influencia de maneira crescente as áreas de cartografia, análise de recursos naturais, transporte, comunicação, energia e planejamento urbano e regional. Um importante aspecto do geoprocessamento é que as informações georreferenciadas têm como característica principal o atributo de localização, ou seja, estão ligadas a uma posição específica do globo terrestre por meio de suas coordenadas (ZAIDAN 2017).

O geoprocessamento representa um conjunto de tecnologias capazes de coletar e tratar informações georreferenciadas. Com o desenvolvimento constante de novas aplicações, as tecnologias associadas ao geoprocessamento fazem cada vez mais parte do nosso dia a dia. Exemplos dessas tecnologias são o Sensoriamento Remoto (SR), o Sistema de Informação Geográfica (SIG) e o Sistema de Posicionamento Global (GPS) (Gomes, 2020).

O SIG é um conjunto manual ou computacional de procedimentos utilizados para armazenar, recuperar, transformar, visualizar e manipular dados referenciados espacialmente num ambiente de respostas a problemas. Trata-se de um banco de dados indexados espacialmente, sobre o qual opera um conjunto de procedimentos para responder a consultas sobre entidades espaciais (ARONOFF, 1989; BURROUGH, 1986; COWEN, 1988; SMITH et al., 1987).

Os primeiros SIGs surgiram na década de 60, no Canadá, como parte de um programa governamental para criar um inventário de recursos naturais. Seu manuseio era muito complexo, pois não existiam monitores gráficos de alta resolução. Os computadores necessários eram excessivamente caros, assim como a mão de obra. A década de 80 representa o momento

quando a tecnologia de sistemas de informação geográfica inicia um período de acelerado crescimento que dura até os dias de hoje (CÂMARA et al., 1996).

O uso e ocupação do solo e problemas ambientais associados são campos muito beneficiados pelas tecnologias SIG. Por exemplo, Machado et al. (2017) relataram que, quando não controlado e não legalizado, o crescimento populacional, juntamente com o aumento desenfreado do número de construções civis, tende a impermeabilizar o solo do local, reduzindo ou extinguindo a capacidade de infiltração do solo. Conseqüentemente, além de alterar o ciclo da água, esse processo causa aumento do escoamento superficial e toda sorte de problemas associados, com particular destaque para o assoreamento e a poluição dos corpos hídricos. Neste contexto, os autores sugerem que programas de geoprocessamento SIG (Sistema de informações Geográficas) são ferramentas que, por meio da utilização de imagens de satélite e outros dados vetoriais georreferenciados, adquiridos por múltiplas fontes, visam gerar informações que deem suporte para gerir essas situações.

Outra aplicação do geoprocessamento é na avaliação de áreas vulneráveis a inundações, pois os mapas de inundação gerados são ferramentas que possibilitam delimitar superfícies expostas a períodos de cheias, e demonstram ser excelentes alternativas na prevenção, controle e gestão das áreas de risco, bem como no planejamento urbano (SILVA e FERREIRA, 2020).

Em estudos epidemiológicos, o uso de ferramentas de geoprocessamento é essencial para o entendimento da dinâmica com que ocorrem os fenômenos epidemiológicos, bem como determinação de áreas de riscos e vulnerabilidade. (SILVA e ZAIA, 2019).

Para Wu et al. (2013), o geoprocessamento não se limita apenas ao planejamento urbano, mas também é de grande valia para os serviços meteorológicos. Segundo os autores, com o uso crescente de serviços de observação e previsão do tempo, há uma necessidade urgente de um sistema eficaz para gerar serviços e produtos de alta qualidade.

### **3.3 Classificação de imagens digitais**

O mapeamento baseado em imagens de satélite fornece uma ferramenta essencial para avaliar os padrões de uso e cobertura da terra atuais e históricos (REN et al., 2019). Além disso, essa ferramenta tornou-se um suporte útil para a investigação de possíveis mudanças no uso da terra em um futuro próximo (REN et al., 2019).

A classificação de uma imagem de satélite consiste na identificação de determinados elementos nelas presentes através da associação de cada pixel ou regiões a uma determinada classe pré-estabelecida (FITZ, 2008). O rápido desenvolvimento de algoritmo de aprendizado

de máquina, permite que a avaliação e classificação sejam quantificadas com mais precisão (RAHMANA et al. 2020).

Segundo Moreira (2003), existem dois métodos mais comuns para a classificação: supervisionados e não supervisionados. A classificação supervisionada envolve o uso de dados de treinamento, onde os pixels da imagem são definidos a partir de descrições numéricas a fim de definir a quantidade e as classes temáticas que serão utilizadas na classificação. Na classificação não supervisionada o processo ocorre sem que o usuário necessite de amostras para treinar o classificador. Neste caso, trata-se de um processo essencialmente realizado pelo computador, em que o usuário apenas especifica o número de classes espectrais desejado e elas são criadas a partir de faixas dos valores de pixel. Assim, os pixels são agrupados de acordo com a similaridade espectral, a limitação neste caso reside no fato de que o usuário terá que interpretar que tipo de cobertura do solo corresponde uma dada classe criada (CRÓSTA, 2002).

Huang et al. (2017) relatam que um método de conjunto multinível, baseados em pixel e objetos, pode explorar os pontos fortes dos classificadores individuais. Do ponto de vista da interpretação de imagens, o método apresentado, integrando regras e recursos múltiplos baseados em pixels e objetos, é eficaz para a classificação de imagens de alta resolução.

### 3.3.1 Classificação baseada a pixel

O processamento com resolução de pixel é capaz de indicar as trajetórias de mudanças na superfície e revelar os padrões de desenvolvimento urbano. Os métodos supervisionados baseados na definição das classes e na coleta dos pixels de treinamento representativos de cada classe realizam cálculos a partir dessas amostras para serem referência nas demais classificações. Alguns métodos estimam a densidade de probabilidade do comportamento dos dados das classes atribuídas (BRANDT. TSO MATHER, P. 2009; ANDREOLA 2009).

A classificação de imagens de sensoriamento remoto por pixel desempenha um papel importante em várias aplicações, como o reconhecimento de culturas, estudos fenológicos e monitoramento de mudanças na cobertura do solo (DIAS, 2020).

### 3.3.2 Classificação baseada em objetos

Para Costa et al. (2017), a classificação de imagem baseada em objetos permite a aplicação de indicadores para o auxílio da classificação de imagens, sendo importante avaliar o desempenho e as formas de aplicações.

A análise baseada em objetos propicia a geração de atributos geométricos e contextuais. Nos objetos mapeados, realiza-se a extração de atributos, isto permite ponderar particularidades geométricas dos alvos até então não considerados nos classificadores tradicionais, tais como a cor, a textura, o formato, a proximidade de segmentos, entre outras particularidades e refinamentos (RUIZ et al. 2017, NETA, 2009).

### 3.4 Aprendizagem de Máquina (*Machine Learning*)

A Aprendizagem de Máquina (*Machine Learning*) é uma área da Inteligência Artificial, é o campo da ciência que fornece ao computador a habilidade de aprender uma determinada tarefa (HOMEM, 2020).

*Machine Learning* é um dos principais pilares dessa nova era do conhecimento, pois permite a extração de informação utilizando dados de forma eficiente e eficaz, alavancando a computação de alto desempenho (FREITAS e SANTANA, 2019).

Fantinel (2020) relata que os computadores em um determinado período eram apenas capazes de processar e efetuar o que lhes eram instruídos a fazer. Porém novas técnicas e ferramentas computacionais foram surgindo devido à crescente tecnologia, possibilitando os computadores a aprender a se auto programar, característica essa do campo *Machine Learning*.

A técnica de *Machine Learning* consiste na identificação automática através de 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).

Métodos de *Machine Learning* podem ajudar a processar uma ampla variedade de conjuntos de dados de sensoriamento remoto e determinar a relação entre elementos, como por exemplo, entre o contínuo de refletância e os recursos de interesse. Através do *Machine Learning* é possível processar diferentes tipos de dados de sensoriamento remoto e investigar suas aplicações, pois demonstra alta capacidade ao combinar dados de sensoriamento remoto e métodos de aprendizado de máquina para mapear diferentes recursos naturais, fornecendo mapas temáticos para a superfície da Terra (SHIRMARD, et. al, 2022).



O *Machine Learning* tem sido considerado confiável na aplicação do sensoriamento remoto, devido sua eficiência e precisão na classificação de imagens de sensoriamento remoto. O *Machine Learning* torna possível gerenciar dados de alta dimensão e mapear recursos de acordo com as características do alvo (Maxwell et al., 2018).

Por ser um domínio muito amplo, o *Machine Learning* se ramificou em vários subcampos que lidam com diferentes tipos de tarefas de aprendizado. Pode-se dividir as tarefas de aprendizagem de acordo com a natureza dessa interação. A primeira distinção a ser observada é a diferença entre aprendizado supervisionado e não supervisionado (SHWARTZ-SHALEV e BEN-DAVID, 2014).

Para Monard e Baranauska (2003), no aprendizado supervisionado é fornecido ao algoritmo de aprendizado, ou *indutor*, um conjunto de exemplos de treinamento para os quais o rótulo da classe associada é conhecido. 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.

O aprendizado supervisionado descreve um cenário no qual a “experiência”, um exemplo de treinamento, contém informações significativas as quais a experiência aprendida deve ser aplicada. No aprendizado não supervisionado, no entanto, não há distinção entre dados de treinamento e de teste. (SHWARTZ-SHALEV e BEN-DAVID, 2014).

O aprendizado não supervisionado tem como objetivo encontrar semelhanças entre os objetos que formam o contexto de informações sem rótulos. Essas aplicações de aprendizagem não supervisionada estão divididas em categorias, tais como aprendizagem hierárquica (intimamente relacionado com o “*deep learning*” e redes neurais multicamadas), “*clustering*” ou agrupamento de dados, modelos de variáveis latentes, redução de dimensionalidade e detecção de “outliers” ou dados discrepantes (USAMA et al., 2019).

#### 3.4.1 Modelos de aprendizagem de *Machine Learning*

A eficácia do *Machine Learning* depende da quantidade e qualidade dos dados disponíveis e dos modelos de aprendizagem. Esses modelos, que são utilizados para representar as aplicações, estão cada vez mais sofisticados e hábeis em resolver problemas complexos graças ao apoio da ciência básica, especialmente a Matemática e a Estatística (FREITAS e SANTANA, 2019).

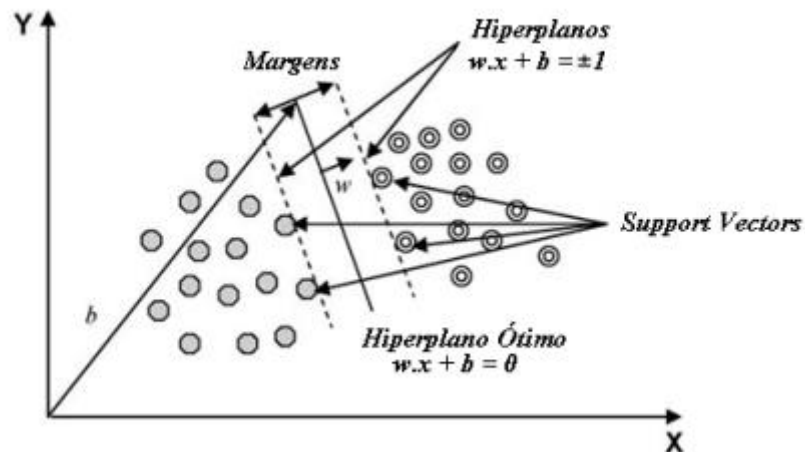
O *Machine Learning* oferece uma vasta gama de opções de algoritmos, com a capacidade de classificar conjuntos de elementos (TAKAKURA et al., 2018). Os algoritmos de *Machine*

*Learning* 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 (CARVALHO 2011). No aprendizado supervisionado estão presentes técnicas como regressão linear, regressão logística, redes neurais artificiais (ANN), máquina de suporte vetorial (SVM), árvores de decisão (DT), k-vizinhos mais próximos (KNN) e Bayes (HONDA et al., 2017).

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

Baseada na Teoria da aprendizagem estatísticas, a *Support Vectors Machine* foi desenvolvida por (VAPNIK, 1995), com o intuito de resolver problemas de classificação de padrões. Esta técnica, originalmente desenvolvida para classificação binária, busca a construção de um hiperplano (Figura 2) como superfície de decisão, de tal forma que a separação entre exemplos seja máxima (GONÇALVES, 2010).

Figura 2 - Representação do SVM.



Fonte: (Kavzoglu e Colkesen, 2009).

O SVM é um dos classificadores clássicos mais populares e classificam os dados tentando criar hiperplanos que melhor separam os dados em suas classes (JUNIOR, 2020).

O algoritmo SVM possui uma teoria bem estabelecida dentro da matemática e estatística, apresenta robustez em dados de grandes dimensões como, por exemplo, imagens de

satélite, possui ótima capacidade de generalização, ou seja, apresenta eficiência para prever dados que não pertençam ao conjunto utilizado no treinamento (SMOLA et al. 2000).

Garofalo et al. (2015) relatam que inúmeros trabalhos têm mostrado que técnicas de classificação não paramétricas, com o emprego de classificadores SVM, têm se destacado na classificação orientada a objetos.

### 3.4.3 Árvores de decisão (*Decision Tree* – DT)

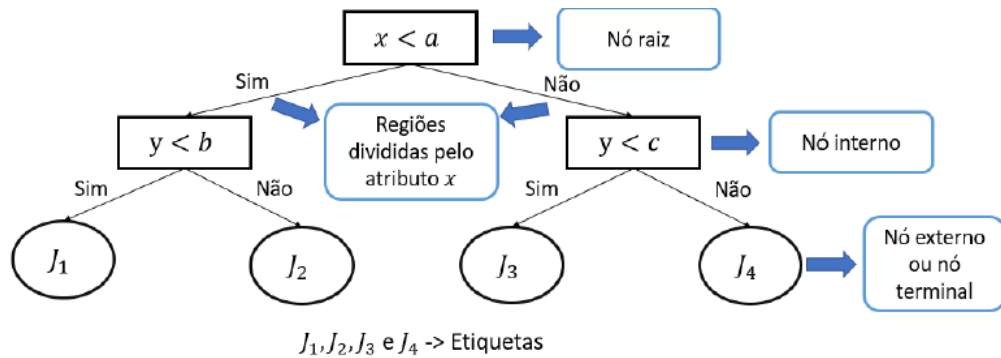
O modelo de classificação *Decision Tree* consiste em um modelo supervisionado capaz de cumprir com tarefas de regressão e de classificação. É um modelo não-paramétrico que modela relações complexas entre as entradas e saídas de um problema de classificação ou regressão, sem a necessidade de assumir hipóteses a *priori* (GUARNIZZO, 2021; ALVARENGA 2018).

Uma árvore de decisão é um preditor, que prediz o rótulo associado a uma instância  $x$  seguindo de um nó raiz de uma árvore para uma folha. Para simplificar as árvores de decisão em cada nó no caminho da raiz para a folha, o filho sucessor é escolhido com base na divisão do espaço de entrada. Normalmente, a divisão é baseada em uma das características de  $x$  ou em um conjunto predefinido de regras de divisão (SHWARTZ-SHALEV e BEN-DAVID, 2014).

Uma árvore de decisão possui uma estrutura composta por nós que são arranjados como em uma ramificação. O nó de decisão ou também chamado nó raiz está sempre no topo da estrutura, e os nós internos contêm um teste de valor sobre um dos atributos mais relevantes. Já os nós-folha, nas extremidades, referem-se às classes da variável resposta e representam o resultado da predição obtida pelo modelo (GOEBEL; GRUENWALD, 1999) (WITTEN; FRANK; HALL, 2011).

O método *Decision tree* (Figura 3) é uma das estruturas de dados mais comuns e poderosas em toda a ciência da computação porque o custo computacional para a construção de uma árvore é mais baixo que outros algoritmos (MARSLAND, 2014).

**Figura 3** - Estrutura Clássica de uma árvore de Decisão



Fonte: Perea, et. al, 2019.

#### 3.4.4 Florestas aleatórias (*Random Forest* – RF)

O método *Random Forest* (RF) consiste em uma técnica de classificação introduzida por Breiman (2001), a qual trabalha com uma grande coleção de árvores de decisão decorrelacionadas. A partir das amostras de treinamento originais são criados diferentes conjuntos amostrais aleatórios, os quais permitem a criação de árvores independentes, onde o resultado da classificação de cada uma das árvores para um determinado objeto é chamado de voto. (IBAÑEZ, 2016).

O RF apresenta aprendizado supervisionado, ou seja, dado as instâncias de treinamento e as etiquetas que indicam a classe de cada uma, o algoritmo aprende a classificá-las sem que haja a intervenção de um usuário no processo. O método vem sendo largamente utilizado nas mais diversas áreas do aprendizado de máquina e vem apresentando ótimos resultados (GUEDES E GUIMARÃES, 2020).

#### 3.4.5 K-Vizinhos mais próximos (*K-Nearest Neighbors* - KNN)

O modelo de classificação *K* Nearest Neighbors (KNN) classifica, para cada instância não rotulada, seus  $k$  vizinhos mais próximos no espaço multidimensional de características abrangendo um conjunto de instâncias de treinamento. A partir dessa classificação, é atribuído um valor de classe de acordo com a maioria de uma classe particular dentro dessa vizinhança (WIELAND e PITTORE, 2014).

Os  $k$  vizinhos mais próximos são identificados por uma medida de distância que compara os vetores de características da instância não rotulada e o conjunto de instâncias de treinamento

fornecidas ao classificador. Definida a lista do vizinho mais próximo, a previsão é baseada em votação (WIELAND e PITTORE, 2014; WU et al. 2007).

O KNN forma uma base generalizada em um conjunto de treinamento, maximizando a acurácia da classificação. O algoritmo presume que o conjunto de treinamento seja composto pelas variáveis descritivas e pela sua classificação, de maneira que o KNN utiliza essas variáveis para classificar um novo item (GAO E GAO, 2010; SU 2011).

#### 3.4.6 *Naive Bayes* – NB

O modelo *Naive Bayes* (NB) consiste em uma classificação probabilística simples que assume a distribuição normal do espaço de características para cada classe. Com base em instâncias de treinamento selecionadas, o algoritmo NB estima vetores médios e matrizes de covariância para cada classe e os utiliza para previsões (BRADSKI et al., 2008; FUKUNAGA 1990).

Para Zhang, (2004), o *Naive Bayes* é um dos mais eficientes e eficazes algoritmos de aprendizado indutivo para aprendizado de máquina e mineração de dados, pois seu desempenho competitivo na classificação é surpreendente. O algoritmo apresenta independência condicional do valor de qualquer outra característica. Trata-se de um conjunto de algoritmos de aprendizado supervisionado baseado na aplicação do Teorema de *Bayes*.

O classificador NB supõe que a presença de uma característica particular em uma classe não está relacionada com a presença de nenhuma outra característica, condição que é chamada de independência condicional (GUARNIZZO, 2021).

### 3.5 **Uso e Cobertura do Solo**

#### 3.5.1 Caracterização do Tema

O uso e cobertura do solo é uma área de estudo que permite avaliar o impacto dos processos naturais e das atividades humanas sobre as características da superfície do planeta e como essas atividades modificam o tipo de cobertura existente ao longo do tempo. Como em geral se refere a estudos comparativos, aspectos como o processo histórico de colonização do local a ser avaliado, o regime político e os aspectos socioeconômicos, urbanos e ambientais devem ser levados em consideração. Esse conjunto de elementos articulados nos permite uma

melhor compreensão do ordenamento territorial e dos problemas ambientais decorrentes das alterações antrópicas e naturais (ALVES e CONCEIÇÃO, 2015).

O tema uso e ocupação do solo reúne informações sobre o grau de preservação, conservação ou artificialização de um determinado lugar do globo terrestre, possibilitando uma avaliação na tomada de decisão, visando a eliminação ou diminuição do fator causador das alterações (ALVES e CONCEIÇÃO, 2015).

O monitoramento de mudanças no uso e ocupação do solo em áreas urbanas fornece informações para suporte à tomada de decisão no planejamento urbano ambiental e, conseqüentemente, na gestão dos recursos naturais (RIEGEL et al., 2019).

Ocupações inadequadas em áreas dos leitos de cursos d'água e margens dos lagos e represas geram problemas ambientais. Maior compactação do solo, diminuição da taxa de infiltração de água no solo, maior processo de erosão dos leitos, poluição do lençol freático e águas superficiais, desmatamento e perda de nutrientes do solo estão entre os principais problemas (DOURADO e PINHEIRO, 2018).

Nesse sentido, o mapeamento do uso e ocupação é essencial para a compreensão dos padrões de ocupação do espaço, o qual por sua vez, tem como principal agente transformador a ação antrópica. Sendo assim, é necessária a constante atualização dos dados de uso e ocupação, o que permite a análise de tendências com o intuito de fornecer informações a tomadas de decisões (ALVES e CONCEIÇÃO, 2015). Portanto, o mapeamento do uso e ocupação do solo é fundamental para o planejamento ambiental e auxiliam na eliminação ou diminuição do fator causador das alterações (SANTOS, 2004; RAMOS et al. 2019).

Por fim, deve ser destacado que o uso e a ocupação do solo resultam de uma mistura de fatores culturais, econômicos e ambientais, a maioria dos quais não pode ser diretamente determinada apenas por técnicas de sensoriamento remoto. Todavia, a técnica oferece informações que auxiliam no processo das análises. Nesse sentido, aspectos físicos da superfície são traduzidos em classes de cobertura do solo como “vegetação rasteira”, “concreto”, “água”, entre outros, possibilitando a criação de mapas temáticos sobre determinada área (JENSEN e COWEN, 1999; BARNSELEY et al., 2001; NOVO, 2017; RIBEIRO, 2014). Portanto, a análise das condições de uso e ocupação do solo começa pelo mapeamento da cobertura do solo e sua evolução ao longo do tempo.

### 3.5.2 Mapeamento da cobertura do solo

Embora o mapeamento da cobertura do solo possa ser realizado por avaliação direta da área de interesse, o processo pode se tornar altamente oneroso, enviesado e inviável no tempo quando se trata de grandes áreas. Nesse sentido, o sensoriamento remoto tem se revelado uma alternativa muito eficaz e mundialmente usada para o mapeamento de regiões de interesse pelas mais diversas áreas do conhecimento (GONG, et al., 2015; HADJIMITSIS, et al. 2010; LIU, et al. 2018). O geoprocessamento se associa ao sensoriamento remoto e permite a classificação digital de uma imagem. Pode-se dizer que o mapeamento da cobertura do solo é uma aplicação importante do geoprocessamento, que oferecendo a possibilidade de separação da superfície em temas, gera mapas temáticos de elementos relacionados às estruturas territoriais, à geografia e aos recursos naturais, por exemplo. O conjunto de informações sistematizadas da área analisada podem facilitar o delineamento de um problema, a tomada de decisões e o planejamento (BARBARA e RIBEIRO, 2019).

Nesse aspecto, vale ressaltar que, por uma questão de resolução espacial dos sensores, os primeiros estudos não encontraram uma aplicação direta nas questões urbanas. Atividades como o monitoramento ambiental e o planejamento urbano estão intrinsecamente relacionados à natureza heterogênea dos espaços urbanos. Entretanto, com o desenvolvimento dos satélites de observação da Terra, houve aprimoramento das órbitas e da tecnologia dos sensores embarcados, o que tornou possível imagens mais precisas e com maior detalhamento dos objetos presentes na superfície. A série de satélites do programa LANDSAT é um dos exemplos de maior sucesso desse aprimoramento (ALMAZROUI, et al., 2017; LI, et. al., 2015; YUAN et al., 2005). Seu acervo histórico e a constante atualização tecnológica dos novos satélites que vão se incorporando à série garantem uma importância ímpar para a compreensão dos grandes problemas ambientais do mundo moderno. Com esse tipo de programa, o sensoriamento remoto ganhou um espaço significativo nos estudos voltados à compreensão do uso e ocupação do solo em áreas urbanas (CAI, et al., 2019; FANG e YU, 2017).

Dentre os mais variados temas abordados por estudos relacionados ao mapeamento da cobertura do solo, a urbanização é aquele de interesse no presente estudo. Ainda assim, o campo de conhecimento é muito amplo e a última década experimentou um avanço considerável nos sensores e nas técnicas matemáticas de classificação da cobertura do solo em áreas urbanas, com destaque para imagens de altíssima resolução fornecidas por satélites modernos, como por exemplo, os das séries WorldView, GeoEye e constelação Plêiades (PANAGIOTAKIS, et al., 2018; HUANG, et al, 2018). Essas imagens têm sido capazes de revelar eficazmente as

verdadeiras informações estruturais da paisagem urbana. A multiplicidade de ângulos utilizada na captura das imagens tem permitido ainda um detalhamento tridimensional dos elementos da superfície, revelando cenas urbanas muito mais complexas do que era possível com as técnicas até recentemente usadas.

É no contexto de elevado refinamento espacial, com resoluções da ordem de um metro ou menos, que o presente estudo se insere, usando o lago Igapó, e toda sua área de drenagem heterogeneamente urbanizada, como objeto de investigação. A investigação pretende fazer uso de todo esse conjunto de ferramentas oferecidas pelo moderno sensoriamento remoto, geoprocessamento e *Machine learning*, para oferecer informações da superfície que ajude a entender de que forma Soluções Baseadas na Natureza (SbN) podem ser eficazmente propostas para problemas complexos que impactam o lago Igapó, decorrente a urbanização.

### **3.6 Soluções Baseadas na Natureza ( SbN)**

O conceito de soluções baseadas na natureza (SbN) surgiu nos anos 2000 e foi ganhando destaque entre organismos internacionais e na União Europeia como uma aposta para a adaptação às mudanças climáticas, pois as SbN são entendidas como soluções que de alguma forma se inspiram, copiam ou tomam como base processos naturais para gerar benefícios sociais, ambientais e econômicos para a sociedade (HERZANG, 2020).

Para Devecchi et. al (2020), as SbN apresentam entre seus objetos os problemas relacionados com o manejo das águas em áreas urbanas, reconhecendo a natureza e seus processos, essencialmente o ciclo hidrológico, como elementos fundamentais na construção do aglomerado urbano. As SbN se propõem a substituir as intervenções humanas poluidoras ou ambientalmente agressivas por práticas ecológicas inspiradas em ecossistemas saudáveis.

As intervenções feitas nas paisagens urbanas, periurbanas e rurais, com alterações na cobertura do solo e nos processos naturais, eliminaram ou alteraram ecossistemas que mantinham as funções ecológicas dessas áreas, o que tem levado, muitas vezes, a consequências dramáticas (FRAGA e SAYOGA, 2020).

Segundo Rodrigues (2021) é possível citar algumas Soluções Baseadas na Natureza que podem ser aplicadas nas cidades e fazer parte dos centros urbanos, auxiliando a amenizar e/ou eliminar as problemáticas ambientais, ao mesmo tempo em que proporcionam diversos outros benefícios para o ser humano e a biodiversidade do planeta.

Os extremos como inundações e alagamentos ou da falta de água, a poluição generalizada do ar, das águas e do solo, com impactos diretos sobre a saúde e o bem-estar das



peças, e as emissões de gases de efeito estufa que aceleram o aquecimento global, têm levado à busca por alternativas mais harmoniosas com a natureza e seus processos e fluxos. Nesse contexto as Soluções baseadas na Natureza (SbN) estão se tornando alternativas cada vez mais aceitas e aplicadas para enfrentar os diversos desafios e manter o funcionamento das cidades (FRAGA e SAYOGA, 2020).

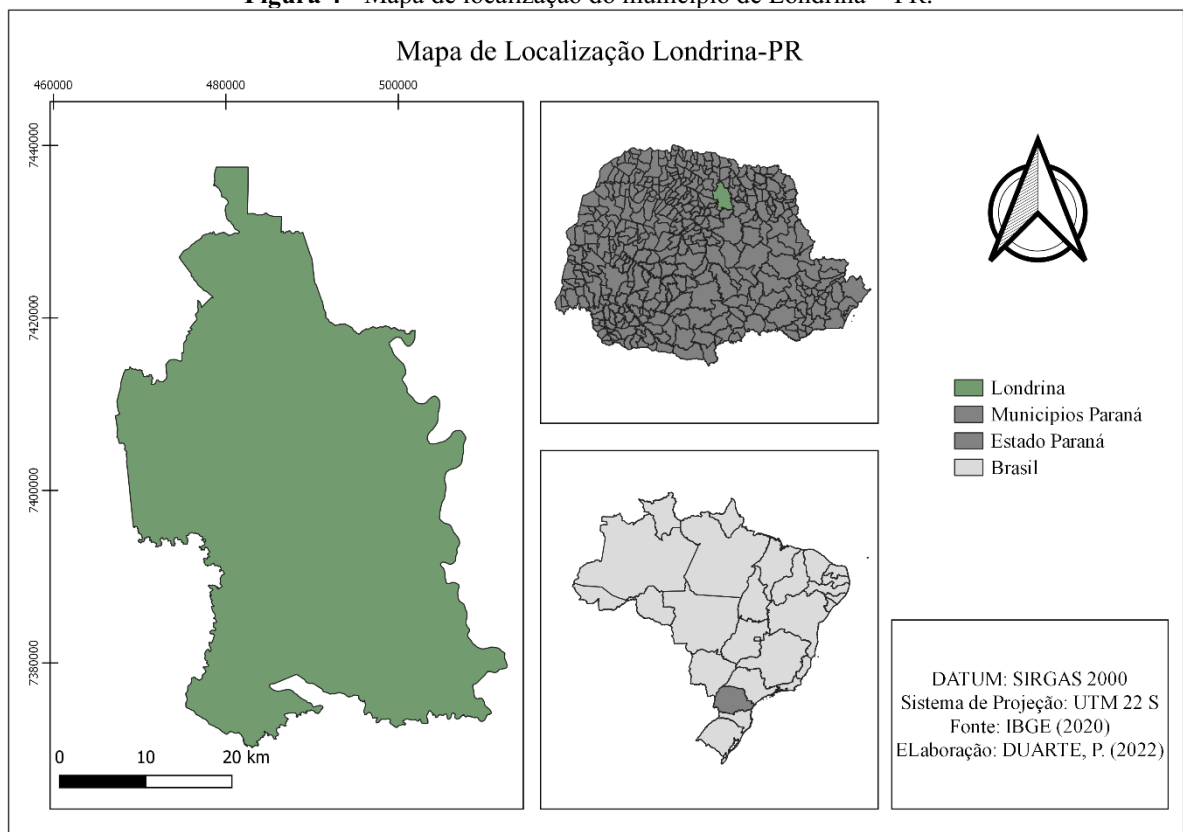
## 4 METODOLOGIA

### 4.1 Caracterização da área de estudo

#### 4.1.1 Área de estudo

O presente estudo foi realizado no município de Londrina-PR, em específico na parte alta da bacia do Ribeirão Cambé em uma área que engloba o Lago Igapó (Figura 4). Londrina é um município localizado no norte do estado do Paraná, cuja criação e instalação ocorreu no ano de 1934 (PREFEITURA DE LONDRINA, 2020). Atualmente o município conta com uma população de 575.377 habitantes segundo a estimativa da população em 2020, sendo a 2ª cidade mais populosa do estado do Paraná. O município possui área territorial de 1651 km<sup>2</sup>.

**Figura 4 -** Mapa de localização do município de Londrina – PR.



Fonte: Da autora, 2022.

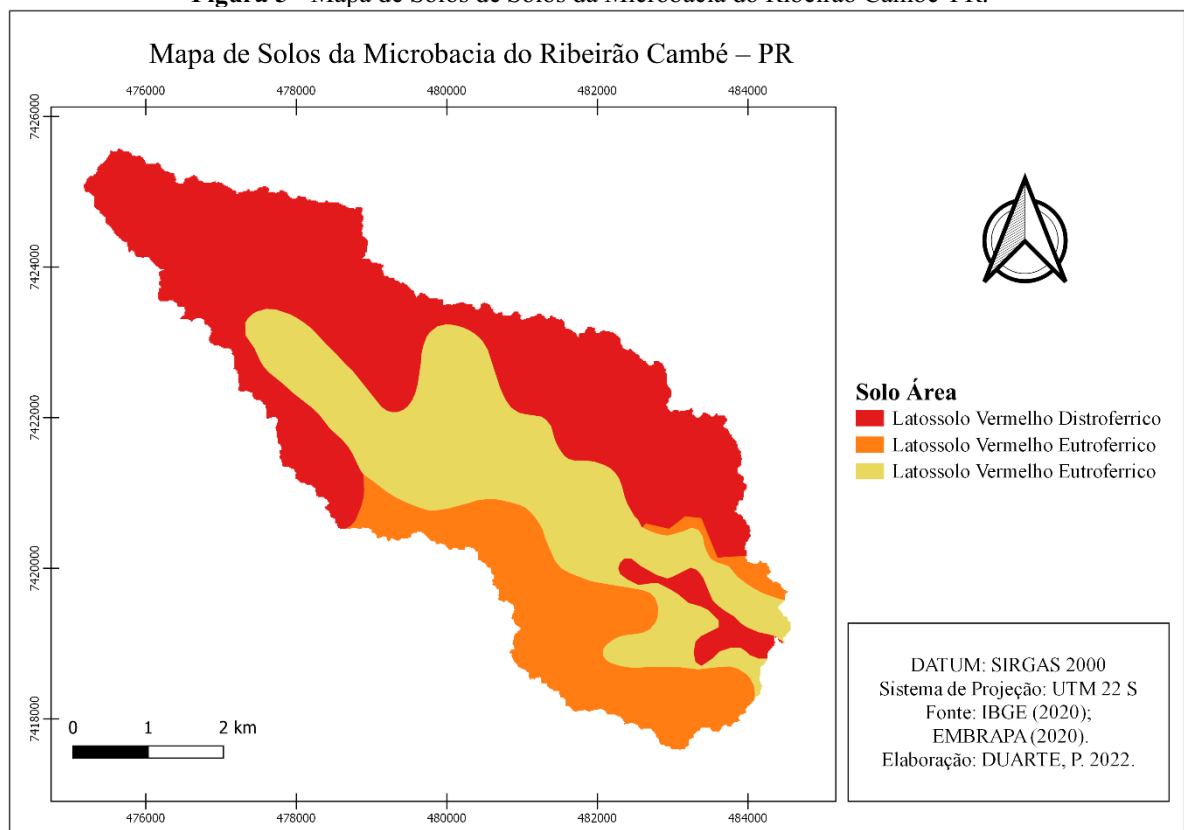
#### 4.1.2 Clima

O clima da região é classificado como temperado úmido com verão quente (Cfa conforme a classificação Köppen e Geiger). A temperatura média é de 20,9 °C e a pluviosidade média anual de 1429 mm (IAP, 2020). Entretanto, a área se encontra numa zona de transição e muito próxima de outros padrões climáticos que variam de clima temperado a inverno seco.

#### 4.1.3 Solos

O município de Londrina, situa-se no Terceiro Planalto paranaense, sobre a formação Serra Geral, os quais são essencialmente constituídas por basalto. O relevo predominante na área é o suave ondulado. Além disso, o tipo de solo predominante é o Nitossolo Vermelho eutroférico (54% da área total do município), seguido por Neossolo Litólico eurófico (26%), Latossolo Vermelho distroférico (10%), Latossolo Vermelho eutroférico (8%) e Nitossolo Vermelho distroférico (2%), (TRABAQUINI et al. 2010). Os solos presentes na microbacia do Ribeirão Cambé são apresentados na figura 5.

**Figura 5** - Mapa de Solos de Solos da Microbacia do Ribeirão Cambé-PR.



Fonte: Da Autora, (2022).

#### 4.1.4 Vegetação

O município tem sua composição de vegetação classificada como Floresta Estacional Semidecidual, restando apenas algumas áreas parcialmente preservadas em bosques e unidades de conservação. Com o processo de ocupação, a vegetação original foi substituída, inicialmente por plantações de café, e posteriormente por áreas urbanas. As áreas verdes representam 13% do total da área urbana de Londrina-PR (BARROS et al., 2008).

#### 4.1.5 Hidrografia

O município de Londrina faz parte da Bacia Hidrográfica do Rio Tibagi a qual está entre as maiores bacias hidrográficas do Estado do Paraná. Esta localiza-se na porção centro-leste do Paraná e possui área de aproximadamente 24.530 km<sup>2</sup>, correspondendo a cerca de 13% da superfície do estado. A Bacia abrange o primeiro, o segundo e o terceiro planalto paranaense, atribuindo à área inúmeros ambientes, o que permite dividi-la em três regiões: alto, médio e baixo Tibagi, onde se localiza o município de Londrina (MEDRI et al., 2002).

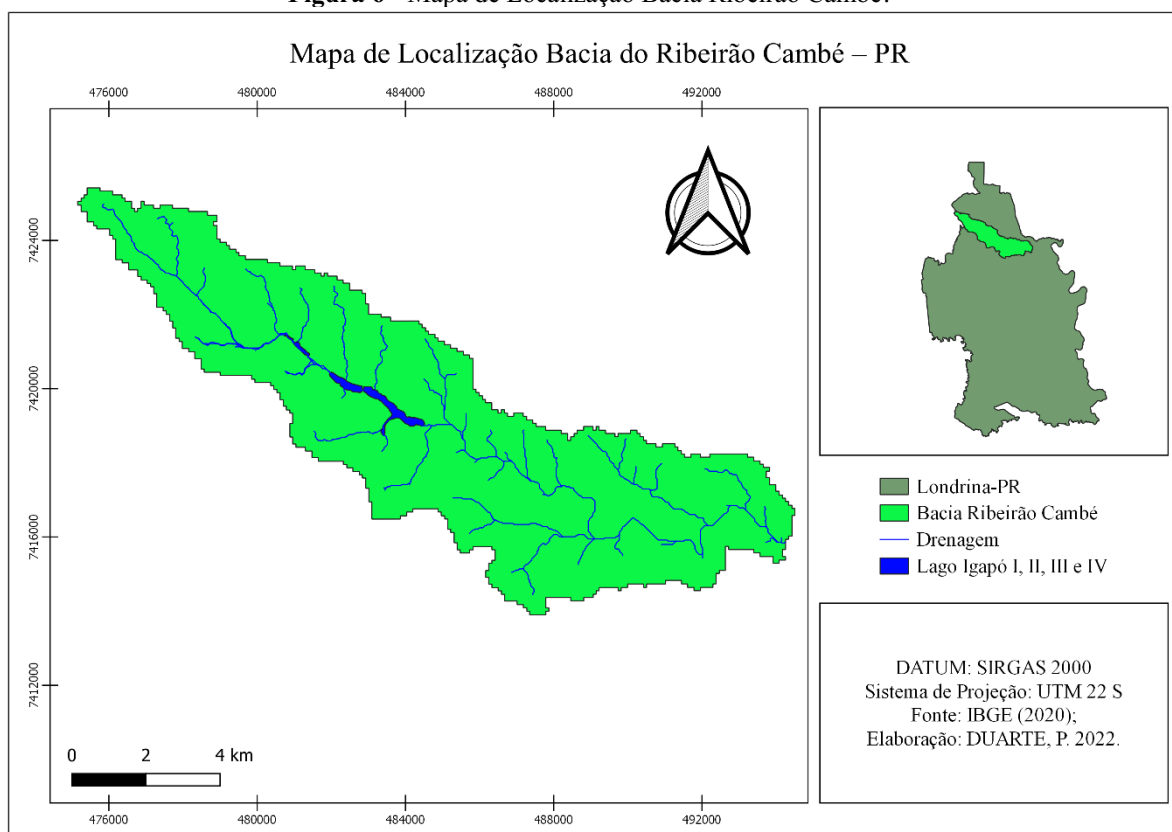
A rede de drenagem do município de Londrina é composta pelas seguintes sub-bacias: Jacutinga, Lindóia, Cambé, Limoeiro, Cafezal e Três Bocas. A hidrografia de Londrina sofreu com a intensa antropização dos mananciais naturais, forçando canalizações de alguns rios (BARROS et al. 2008 e CARDOSO, 2013).

O presente estudo foi realizado nos limites da Bacia do Ribeirão Cambé (Figura 6), o qual vem sofrendo alterações decorrentes da ação antrópica, devido ao desenvolvimento urbano.

CHANG et. al, 2021, relata que a bacia do Ribeirão Cambé, no município de Londrina/PR é pouco ramificada, tem área de 29,95 km<sup>2</sup>, confirmando a área do presente estudo. O perímetro é de 27,14 km e o comprimento do canal principal é de 11,25 km, com uma rede de drenagem total de 24,09 km. A densidade de drenagem é de 0,80 km/km<sup>2</sup>, indicando que a bacia possui uma boa capacidade de drenagem. Este é um importante fator, pois indica o grau de desenvolvimento do sistema de drenagem de uma bacia.

A bacia em estudo nasce no município de Cambé-PR, a oeste do município de Londrina-PR, possui uma área de 77,20 km<sup>2</sup> (Figura 6) e o comprimento do principal curso d'água, o Ribeirão Cambé, é de 25 km. A área da Bacia do Ribeirão Cambé, além do curso d'água principal, conta com tributários como o córrego Capivara, Leme, Água Fresca, Colina Verde, Rubi, Baroré e córrego da Mata, (ARAUJO, 2004 e IPUL, 2013) (Figura 7).

**Figura 6 - Mapa de Localização Bacia Ribeirão Cambé.**



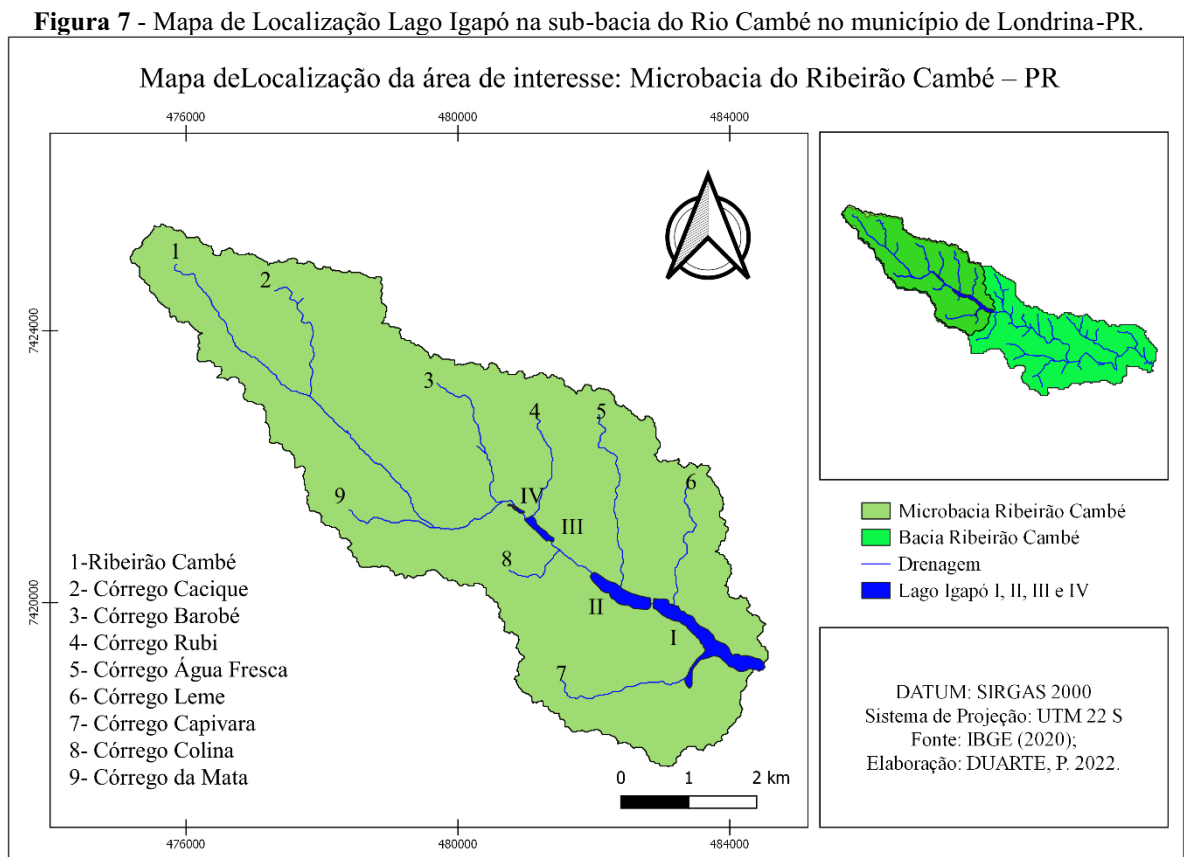
O Ribeirão Cambé corta a cidade de Londrina-PR no sentido noroeste a sudeste e é um dos principais cursos d'água que fazem parte do sistema de drenagem do município. Neste, localiza-se o conjunto de lagos denominados lago Igapó (I, II, III e IV). O Lago (Figura 7) foi inaugurado em 1957 e a obra foi realizada na gestão do prefeito Antônio Fernandes Sobrinho. Posteriormente veio a construção da barragem do lago, causando uma grande desvalorização na região ao seu entorno. A valorização da área se deu apenas na década de 1970, quando o prefeito Dalton Paranaguá decidiu revitalizar o lago para melhorar a saúde e qualidade de vida da população. Para isto, o então o prefeito contratou o paisagista paulistano Burle Marx para desenvolver o projeto de revitalização do Lago Igapó, tornando o local propício para práticas esportivas e de lazer, resultando nas primeiras ocupações próximas ao lago (JUNIOR (2020) A.P.A; NETO, P.C.O, 2020 e LORENZO, 2011).

#### 4.1.6 Lago Igapó

A construção do Lago Igapó não considerou as matas ciliares que são fundamentais para aumentar a infiltração de águas pluviais no solo e protegê-lo contra riscos geomorfológicos, como erosão e o assoreamento subsequente (BORTOLO, 2011; GALHOTE, 2019 e SANTOS et al. 2017).

Porém, o processo de assoreamento do lago Igapó pode estar relacionado aos processos erosivos identificados nos seus afluentes, em suas margens ou em solos adjacentes, principalmente (LORENZO, 2011).

Atualmente, o lago pode ser considerado um exemplo de valorização de espaço público, visto que consiste em uma área de lazer que se caracteriza por uso intenso e diversificado em relação aos segmentos sociais e faixas etárias de seus frequentadores.



Fonte: Da autora, 2022.

## 4.2 Material

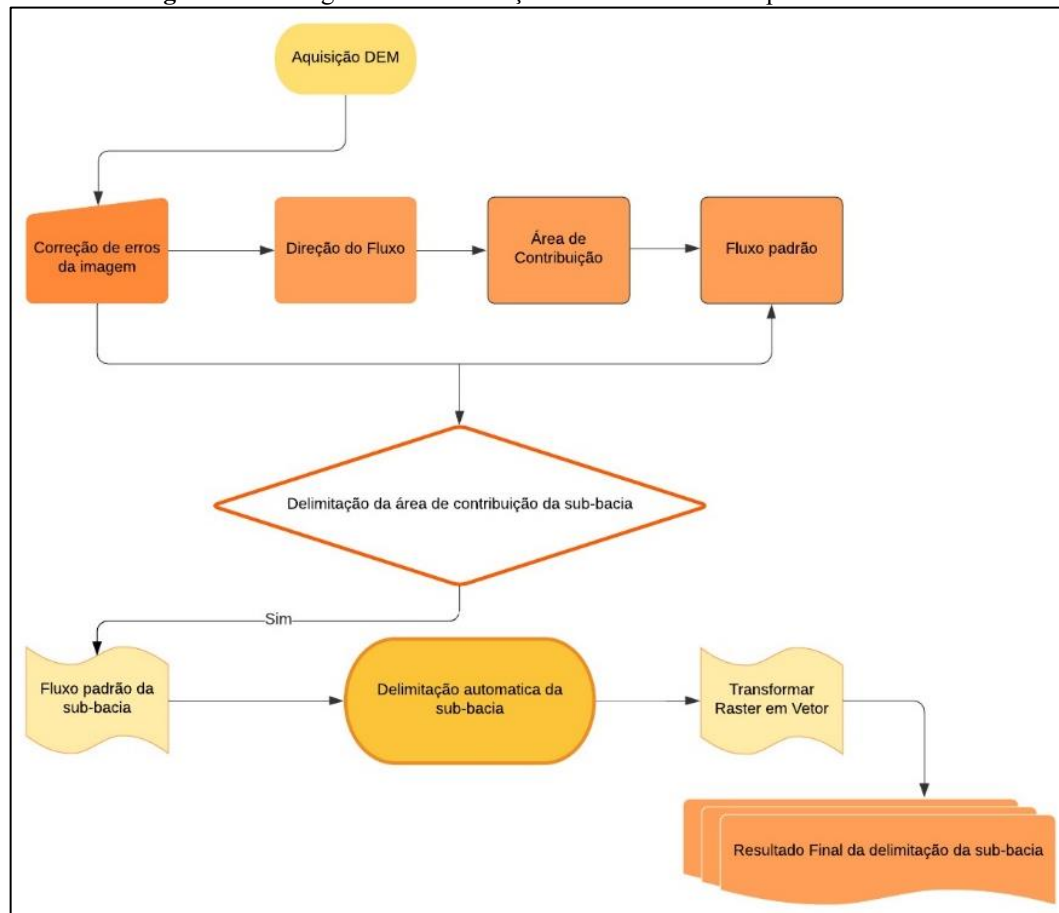
No desenvolvimento da presente pesquisa utilizou-se uma imagem do satélite Pléiades, com altíssima resolução espacial (0,5 metros). Para o processamento dos dados utilizou-se o *software* QGIS versão 3.14 Madeira, além de complementos como o Taudem para delimitação da sub-bacia, o Dzsaka para classificação supervisionada por Píxel, Orpheutoobox para classificação supervisionada por objeto (OBIA), o SCP para realizar a acurácia das classificações, além do *software* Terraview, utilizado para realizar a segmentação, processo que antecede a classificação.

O *Quantum Geographical Information System* (QGIS) é um software livre com código-fonte aberto, multiplataforma de Sistema de Informação Geográfica (SIG) que permite a visualização, edição e análise de dados georreferenciados (QGIS, 2022).

O QGIS permite ao usuário criar mapas com várias camadas usando diferentes projeções, os mapas podem ser montados em diferentes formatos e para diferentes usos, pois o *software* permite compor mapas a partir de camadas raster e/ou vetoriais. Os dados podem ser armazenados como pontos, linhas ou polígonos disponibiliza um número de funcionalidades em constante crescimento através das funções nativas e de complementos (SCHORN & YAMAMOTO, 2021).

## 4.3 Delimitação da sub bacia do Ribeirão Cambé

Para realizar a delimitação da microbacia (Figura 8) foi necessário usar um Modelo Digital de Elevação (MDE). Existem vários modelos atualmente em uso pela comunidade científica e que estão disponíveis em domínio público. Entre esses modelos, destacam-se o ASTER (*Advanced Spaceborne Thermal Emission Radiometer Stereoscopia*) e o SRTM (*Shuttle Radar Topography Mission*). Embora exista uma variedade de modelos e versões, a escolha de um deles depende das características do projeto onde será usado. O MDE ALOS-PALSAR foi escolhido para a delimitação da bacia, devido à alta resolução espacial de (12,5 m) e sua gratuidade.

**Figura 8** - Fluxograma de delimitação de bacia através de ponto exultório.

Fonte: Da Autora, (2022).

#### 4.4 Aquisição de imagem de alta resolução

Com base na definição da área de interesse foi realizada a aquisição de dados do satélite Pléiades junto a empresa Tecterra Geotecnologias e Meio Ambiente LTDA. As especificações da imagem obtida são apresentadas na Tabela 1:

**Tabela 1:** Especificações técnicas da imagem de satélite Pléiades da área de interesse do presente projeto

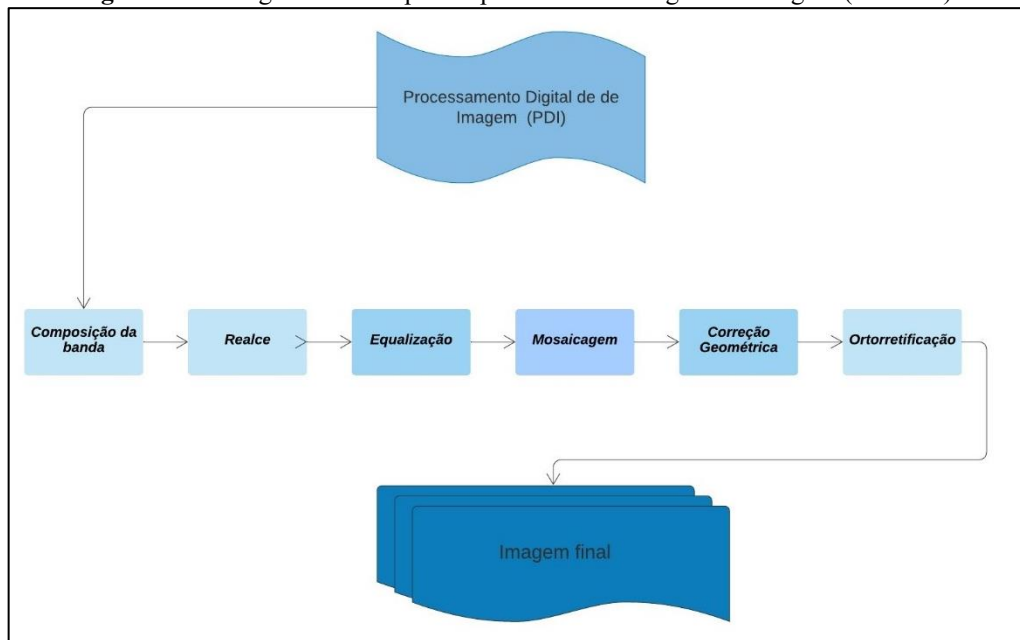
Dados	Especificações
Satélite	Pléiades
Resolução Espacial	2M RGB / 0,5 PAN
Bandas Espectrais	RGB e NIR
Data da cena	16/06/2020
Cobertura de nuvens	0%
Formato	GeoTIFF

Fonte: Da Autora (2022).



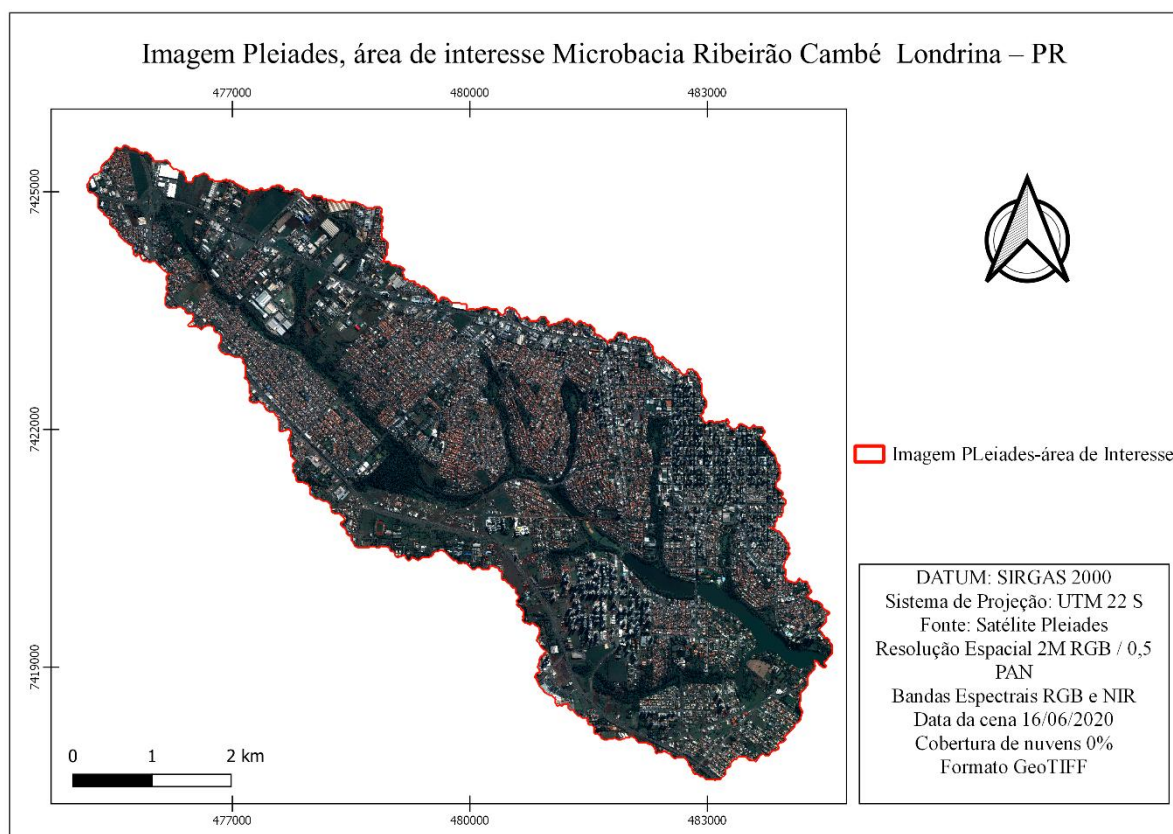
O processamento digital de imagens (PDI) compõe uma parte do Sensoriamento Remoto em que são aplicados determinados processos de ajuste, objetivando a melhoria da qualidade de dados, como a visualização dos alvos e precisão das imagens e, por conseguinte, das informações que serão retiradas delas. Isto é necessário, pois as imagens adquiridas pelos sensores necessitam de correções que os satélites não conseguem realizar (TECTERRA, 2021). A Figura 9 apresenta as etapas do processamento digital realizado na imagem. Com base nas técnicas de pré-processamento aplicadas, que incluem técnicas de composição de bandas, realce, equalização, mosaicagem, correção geométrica e ortorretificação da imagem, obteve-se a imagem ajustada para posterior aplicação dos métodos de classificação. A imagem foi adquirida com todo processo de PDI realizado pelo fornecedor (Figura 10). As etapas envolvendo as técnicas de pré-processamento de imagens podem ser observadas na Figura 9.

**Figura 9** - Fluxograma das etapas do processamento digital de imagem (Pléiades).



Fonte: Do autor(a), 2022.

**Figura 10** - Imagem de alta resolução com área delimitada da área de estudo em Londrina-PR, no ano de 2020.



Fonte: Do autor(a), 2022, adaptado Pléiades (2020).

O processo Digital de Imagem foi realizado pela empresa fornecedora da imagem Pléiades, que de acordo (TECTERRA, 2020), a primeira etapa do PDI contou com a composição das bandas multiespectrais, que contempla a unificação das “4” bandas monocromáticas (bandas referentes aos comprimentos de onda R - vermelho, G – verde, B – azul e NIR - infravermelho próximo, em uma única imagem colorida. O Procedimento de realce é necessário para um melhor aproveitamento na visualização dos objetos geográficos contidos em uma imagem. O processo de equalização visa minimizar feições que podem apresentar tonalidades heterogêneas. A mosaicagem é necessária quando houver várias cenas diferentes compondo uma única área de trabalho, pois algumas imagens podem possuir angulação diferente. A junção destas imagens também tem o objetivo de reduzir o efeito de corte ou emenda entre as imagens. Correções Geométricas são realizadas para adequar o georreferenciamento nativo das imagens a partir de pontos coletados em campo ou de bases cartográficas disponíveis. A ortorretificação objetiva corrigir as deformações geométricas sofridas pelas imagens devido à plataforma do sensor, do ângulo de aquisição, do relevo, da curvatura e rotação da Terra, sendo possível realizar o procedimento através de bases planialtiméticas pré-existentes (TECTERRA, 2020).

## 4.5 Classificação da imagem

A classificação supervisionada é um método que exige que o usuário conheça as características de locais específicos dentro da área a ser classificada e coincidindo com a época de aquisição da imagem. Esse tipo de observação é conhecido como verdade terrestre "*ground truth*" (CRÓSTA, 2002).

### 4.5.1 Segmentação de da imagem

O primeiro passo para realizar a classificação foi a segmentação utilizando o método crescimento de regiões, através do software TerraView. Foram realizados alguns testes (Tabela 2) em relação aos parâmetros de segmentação da imagem a partir do algoritmo de crescimento de região, apontando como melhor resultado o valor 80 de tamanho mínimo dos segmentos e 0,075 para o limiar de similaridade. O critério de similaridade baseia-se em um teste de hipótese estatístico que testa a média entre as regiões. Posteriormente, divide-se a imagem em um conjunto de sub-imagens e então realiza-se a união entre elas, segundo um limiar de agregação definido (INPE, 2006).

**Tabela 2:** Valores de teste para segmentação de imagem no TerraView.

	<b>Tamanho Mínimo do Segmento</b>	<b>Limiar de Similaridade</b>
1	1	0.025
2	10	0.015
3	20	0.035
4	40	0.05
5	60	0.075
6	80	0.075

Fonte: Do Autor(a), (2022).

### 4.5.2 Extração de atributos

Para o processo de extração de atributos criou-se um modelo gráfico dentro do OTB para calcular em conjunto a estatística das bandas 1,2,3 e 4 da imagem Pléiades, através do *ZonalStatistic*. Este processo consiste no cálculo das estatísticas associadas aos segmentos (Tabela, 3), onde se calcula número de pixels, valor mínimo, valor máximo, valor médio, desvio

padrão e variância, para cada segmento e para cada banda. A tabela 3 apresenta a descrição de cada variável.

O algoritmo calcula as estatísticas a partir dos valores do raster, tendo como referência a delimitação dos segmentos da camada vetorial obtida na segmentação. Todavia, é fundamental que os segmentos representem adequadamente os objetos da imagem (GONÇALVES, 2021).

**Tabela 3:** Campos Computados pelo Algoritmo ZonalStatistic (OTB).

<b>Atributo</b>	<b>Descrição</b>
Número de pixels	Número de pixels em cada segmentos
Valor mínimo	Valor mínimo do nível de cinza dos pixels dentro do objeto
Valor máximo	Valor máximo do nível de cinza dos pixels dentro do objeto
Valor médio	Valor médio do nível de cinza dos pixels dentro do objeto
Desvio padrão	Valor do desvio padrão do nível de cinza dos pixels dentro do objeto
Variância	Valor da variância do nível de cinza dos pixels dentro do objeto




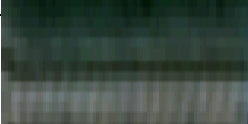




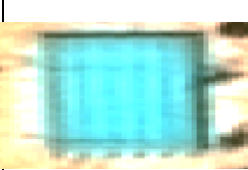

Fonte: OTB *Development Team* (2022), adaptado.

#### 4.5.3 Definição das Classes de uso e ocupação do solo e seleção de amostras de treinamento e validação

Um dos objetivos específicos deste trabalho é a produção de um mapa das condições atuais de uso e ocupação do solo, por meio de imagem de alta resolução espacial. O rápido e intenso processo de urbanização, comum à maioria dos municípios brasileiros, acarreta uma elevada degradação dos recursos hídricos e ecológicos. Nesse sentido, a avaliação do uso e ocupação do solo é uma etapa essencial para o planejamento urbano conservacionista e sustentável. A implantação de soluções baseadas na natureza é altamente dependente do conhecimento pormenorizado das condições atuais de ocupação da área em estudo. Essa avaliação permite um diagnóstico das condições atuais de distribuição no espaço ocupado das áreas naturais e antropizadas.

Para o presente estudo foram estabelecidas classes temáticas específicas para classificação. O Quadro 1 mostra as classes das condições de uso e ocupação adotadas.

**Quadro 1:** Classes de elementos para determinação de uso e ocupação do solo

Classes	Rotulo Subclasse	Subclasses	Descrição	Amostra
Floresta	1	Florestas nativas ou plantada	Áreas contínuas e cobertas por vegetação nativa, ainda que recuperadas, mas com dossel denso e característico da cobertura original da área.	
	1	Árvores de calçamento	Compreende espécies individuais que ocupam áreas de calçamento adjacentes às ruas e avenidas.	
Gramínea	5	Gramado	Inclui áreas ocupadas por gramados de jardins e obras de paisagismo em geral e pastagem.	
Edificação	4	Asfalto	Todo tipo de superfície coberta por asfalto, incluindo ruas, avenidas, ciclovias e estacionamentos.	
	8	Calçada	Qualquer tipo de calçamento a base de cerâmica ou cimento.	
	12	Telhado cinza	Superfícies cobertas por lajes em concreto e telhas de cimento ou fibrocimento.	
	09 10	Telhado cerâmica clara e escura	Qualquer tipo de superfície coberta por telhas de barro, (cerâmica clara e escura).	
	11	Telhado metálico	Telhados de zinco ou qualquer outro tipo de superfície metálica.	
Água	2	Rio	Cursos de água em geral, incluindo córregos e ribeirões.	
	2	Lago	Qualquer tipo de água represada ou que não demonstre orientação definida em seu movimento.	
	3	Piscina	Piscinas de qualquer natureza.	
Outros	7		Sombra	

Fonte: Da Autora, (2022).

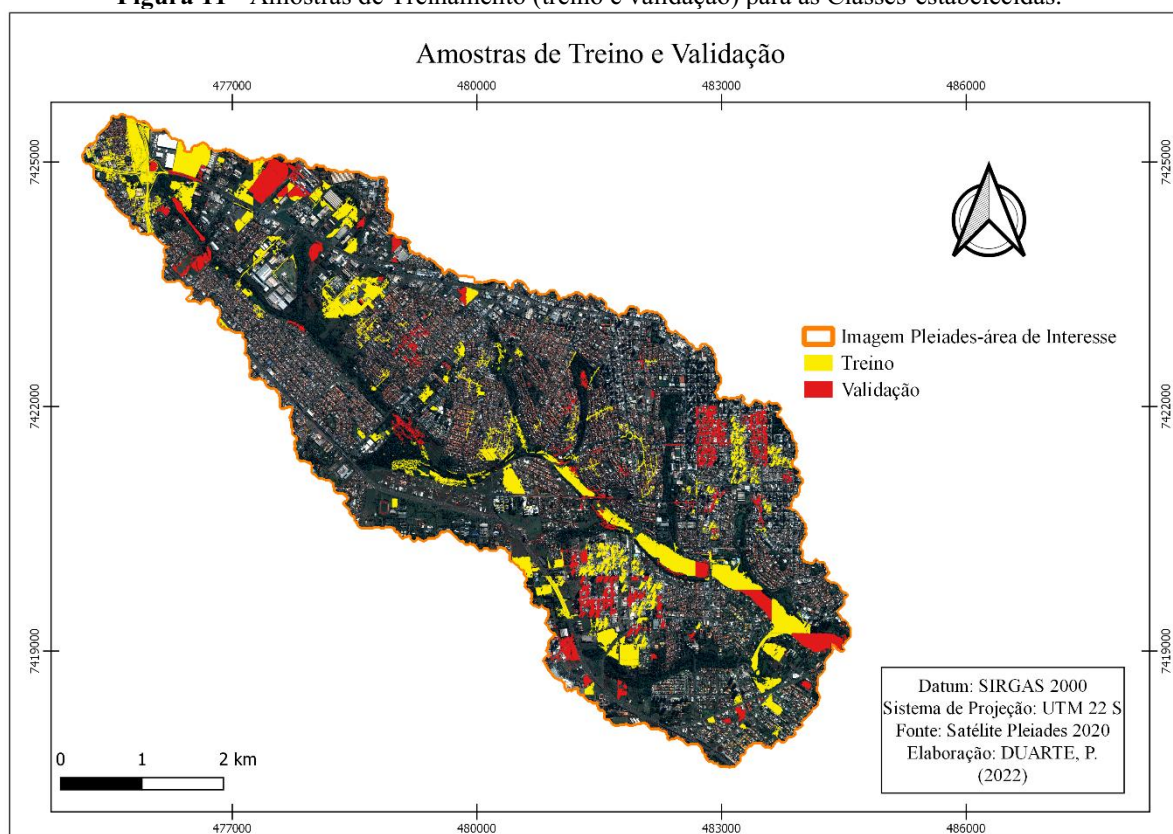
As classes propostas representam os elementos essenciais da paisagem urbana, aprimoradas ao longo do trabalho, principalmente a partir das visitas a campo na área de estudo. Com as classes definidas, partiu-se para o processo de coleta de amostras de treinamento.

As áreas escolhidas para a coleta dos objetos representativos do tipo de cobertura do solo são denominadas de amostras de treinamento. Essas áreas servem para conduzir o algoritmo de classificação na associação dos pixels às correspondentes classes de interesse, sendo um dos componentes mais decisivos na classificação supervisionada. Para coletar um volume de dados de treinamento de forma eficiente são necessários serem adotados alguns critérios, importante não só a quantidade de amostras, mas também a qualidade e a distribuição das amostras (MOBLEY, 2016; RUDKE, 2018).

A seleção das amostras consiste na indicação manual de um número mínimo de segmentos representativos para cada classe definida. As amostras foram selecionadas de forma aleatória, após realizar uma análise visual detalhada da imagem, selecionando segmentos da imagem, distribuídos de maneira uniforme em todas as classes definidas para a área de estudo.

No total foram coletadas 578 amostras (Figura 11), sendo 50 amostras para as subclasses: floresta nativa, plantas e árvores viárias; gramado; asfalto, calçada, telhado cinza, telhado de cerâmica escura e cerâmica clara, telha metálica dentro da classe edificações; para a classe água, subclasse (lagos e rios) foram coletadas 28 amostras de treinamento, devido à baixa presença desses elementos na área de interesse, 50 amostras para piscina e 50 para subclasse sombra.

**Figura 11** - Amostras de Treinamento (treino e validação) para as Classes estabelecidas.



Fonte: Da Autora, 2022.

#### 4.5.4 Treinamento e validação

O processo de treinamento e validação foi realizado através da aplicação TrainVectorClassifier, dentro do OrfeoToolbox(OTB)/Qgis. Os dados de entrada necessários para esta etapa são os vetores contendo os polígonos de treinamento, sendo 70% do total das 578 amostras, e validação, 30% das amostras coletadas na etapa anterior. A figura 12 apresenta as amostras de treino e validação.

O TrainVectorClassifier treina um classificador com base em geometrias, tendo como Output um arquivo de texto, cujo formato e conteúdo correspondem ao tipo de modelo de *Machine Learning* escolhido (OTB,2021).

Para o presente estudo foram escolhidos cinco algoritmos de *Machine Learning*, tendo como saída um modelo de classificação os algoritmos escolhidos foram de classificação supervisionada, sendo eles: *Decision Tree* (DT), *Randon Forest* (RT), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN) e *Normal Bayes*, cada modelo gerado nesta etapa se encontra no Apêndice I a IV.

#### 4.5.5 Classificação

A classificação foi realizada através da função *VectorClassifier* disponível no pacote OTB a partir da Caixa de Ferramentas de processamento do QGIS. Este algoritmo depende do modelo gerado para cada algoritmo, obtido na etapa anterior e dos cálculos das estatísticas obtidas no processo utilizando o *Zonal Statistic* e o vetor da segmentação.

#### 4.5.6 Avaliação dos modelos

O processo de avaliação foi realizado através o plugin SCP no Qgis. Apenas o parâmetro de acurácia conhecido como *F1score* foi calculado de acordo com a metodologia utilizada por Kugler et. al 2018.

A primeira fase para avaliar os dados consistiu na geração da matriz de confusão, a qual representa a confusão entre as classes, conforme mostra o Quadro 2. Existem quatro possíveis saídas da classificação para cada instância:

- Verdadeiro Positivo (TP): corresponde aos dados positivos classificados corretamente pelo algoritmo.
- Falso Negativo (FN): o número de dados positivos classificados incorretamente.
- Verdadeiro Negativo (TN): quando a instância negativa é classificada como tal.
- Falso Positivo (FP): no caso de dados positivos classificação incorretamente.

**Quadro 2:** Matriz de confusão classificação DT.

Classe	1	2	3	4	5	6	7	8	9	10	11	12	Total
1	111646	0	0	0	21019	9394	759	0	0	0	0	0	142818
2	6960	490480	0	0	0	5297	0	0	0	0	0	0	502737
3	0	0	1868	0	0	0	0	41	0	0	0	0	1909
4	0	0	0	393416	0	32919	0	0	0	0	0	0	426335
5	0	0	0	28773	657779	63692	0	0	0	0	0	0	750244
6	0	1148	0	4076	12879	504426	0	0	0	0	0	193	522722
7	260	0	0	0	0	0	359438	0	0	0	0	0	359698
8	0	0	0	0	0	0	0	1783	0	86	24	381	2274
9	0	0	0	0	0	0	0	92	1903	245	250	64	2554
10	0	0	0	0	0	0	0	0	0	2228	0	0	2228
11	0	0	0	0	0	0	0	0	1259	312	10043	280	11894
12	0	0	165	0	82	180	0	0	0	0	0	6434	6861
Total	118866	491628	2033	426265	691759	615908	360197	1916	3162	2871	10317	7352	2732274

- Diagonal Principal: Valores de pixel classificados corretamente

Fonte: Da Autora, 2022.



Guarnizzo (2021) relata que a diagonal da matriz de confusão contém as decisões corretas do classificador, enquanto outros elementos da tabela representam o número de classificações incorretas. A matriz de confusão é a fonte para o cálculo adicional das métricas de sucesso dos algoritmos, incluindo a taxa de verdadeiros positivos, a qual também é chamada de sensibilidade ou *recall*, e a precisão.

Sensibilidade ou *recall*:

Essa métrica avalia a capacidade do método de detectar com sucesso resultados classificados como positivos. Ela pode ser obtida pela Equação (1):

$$\text{Sensibilidade} = \frac{TP}{TP+FN} \quad (1)$$

Precisão:

A precisão é uma métrica que avalia a quantidade de verdadeiros positivos sobre a soma de todos os valores positivos e pode ser obtida pela Equação (2):

$$\text{Precisão} = \frac{TP}{TP+FP} \quad (2)$$

F1-score:

F-score ou *score*  $F_1$  é uma média harmônica calculada com base na precisão e na sensibilidade. Ela pode ser obtida com base na Equação (3):

$$f1 = 2 * \frac{\text{precisão} * \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}} \quad (3)$$

O índice Kappa (tabela 4) é um dos principais métodos utilizados para avaliar a concordância entre a verdade terrestre e o mapa temático. A grande vantagem do uso da estatística Kappa é que, no cálculo do coeficiente Kappa, incluem-se todos os elementos da matriz de confusão (GASPARIN et al., 2013). A matriz de confusão é formada por um arranjo quadrado de números dispostos em linhas e colunas que expressam o número de unidades de amostras de uma categoria particular relativa – inferida por um classificador (ou regra de decisão), comparado com a categoria atual verificada no campo (CONGALTON, 1991).

**Tabela 4:** Grau de exatidão da classificação de acordo com o valor do Índice Kappa.

Valor do Kappa	Força da Concordância
< 0,2	Pobre
0,21 - 0,40	Fraca
0,41 - 0,60	Moderada
0,61 - 0,80	Boa
0,81 - 1,00	Muito boa

Fonte: Adaptado de Bishop, Fienberg and Holland (1975)

A estatística Kappa varia de 0 a 1,0, sendo que o valor de Kappa igual a 0,0 sugere que a classificação não é melhor do que uma classificação aleatória dos pixels (Tabela, 4). No entanto, um valor próximo de 1,0 sugere um processo de classificação altamente eficiente (MOREIRA, 2001).

As métricas (*Recall*, Precisão, *F1-score* e Kappa) utilizadas para a avaliação dos classificadores foram aplicadas para todas as 12 classes utilizadas na classificação. Para a avaliação global de cada classificador as métricas foram a matriz de confusão, acurácia global, Kappa e *F1-score*. Tais valores foram obtidos através do processamento da acurácia do plugin SCP, exceto o valor do *F1-score*.

#### **4.6 Confirmação da exatidão do classificador *Streetview***

Uma forma de confirmar a exatidão do processamento de dados aplicados na presente pesquisa foi através do Google *StreetView*. Através das coordenadas dos alvos classificados foi possível navegar na plataforma para confirmar a precisão e acerto da classificação (Figura 12). Com a classificação pronta, foi realizada a seleção aleatória de amostras classificadas, através da ferramenta de vetor seleção aleatória no Qgis. Posteriormente se extraiu as coordenadas e o shape foi aberto no Google Earth Pro, onde se navegou no *StreetView* confirmando a classe classificada pelo algoritmo.

Figura 12 - Uso do Stretview para confirmação da classificação.



Fonte: Google Earth Pro, 2022.

## 5 RESULTADOS E DISCUSSÃO

Através do processo de segmentação extraiu-se 231.813 objetos dentro dos 30 km<sup>2</sup> da área de estudo. A Figura 13 apresenta a prévia do resultado obtido. Em seguida, iniciou-se a extração dos atributos da imagem já segmentada.

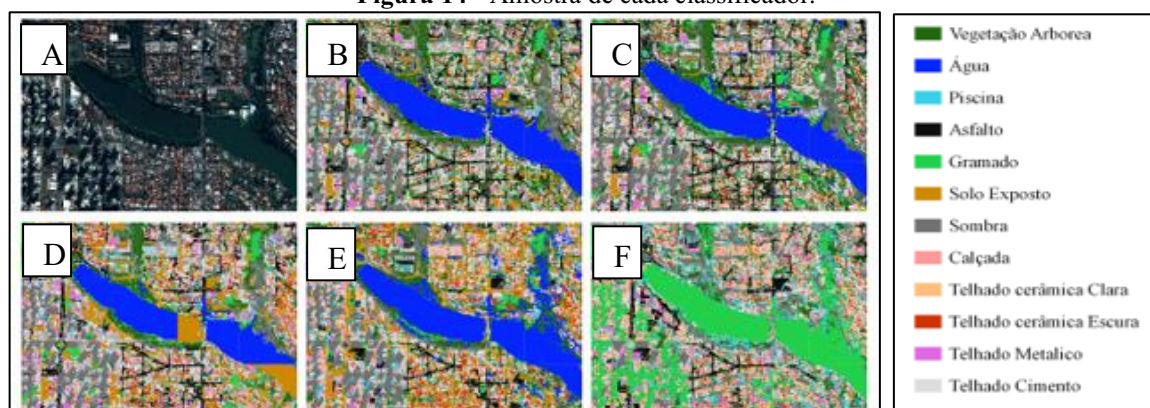
**Figura 13** - Segmentação de imagem no TerraView.



Fonte: Do Autor(a), 2022. A- Imagem Pléiades (2020); B- Imagem Segmentada TerraView

Com a imagem segmentada e o processo de coleta de amostras realizado, iniciou-se o processo de classificação. Esse processo permite que se obtenha a classificação, ou seja, o mapa de uso e ocupação do solo. A Figura 14 apresenta a prévia de cada classificador utilizado. Porém, para decidir qual o melhor classificador foi preciso realizar a etapa de acurácia. Foram obtidos resultados para os cinco classificadores utilizados, sendo eles *Decision Tree* (DT), *Randon Forest* (RT), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN) e *Naive Bayes*. Cada mapa gerado através dos modelos de classificação consta nos Apêndices I a V.

**Figura 14 - Amostra de cada classificador.**



A-Imagem Pléiades ; B- Decision Tree; C- Randon Forest; D- Naive Bayes E- SVM F-KNN

Fonte: Da Autora, 2022.

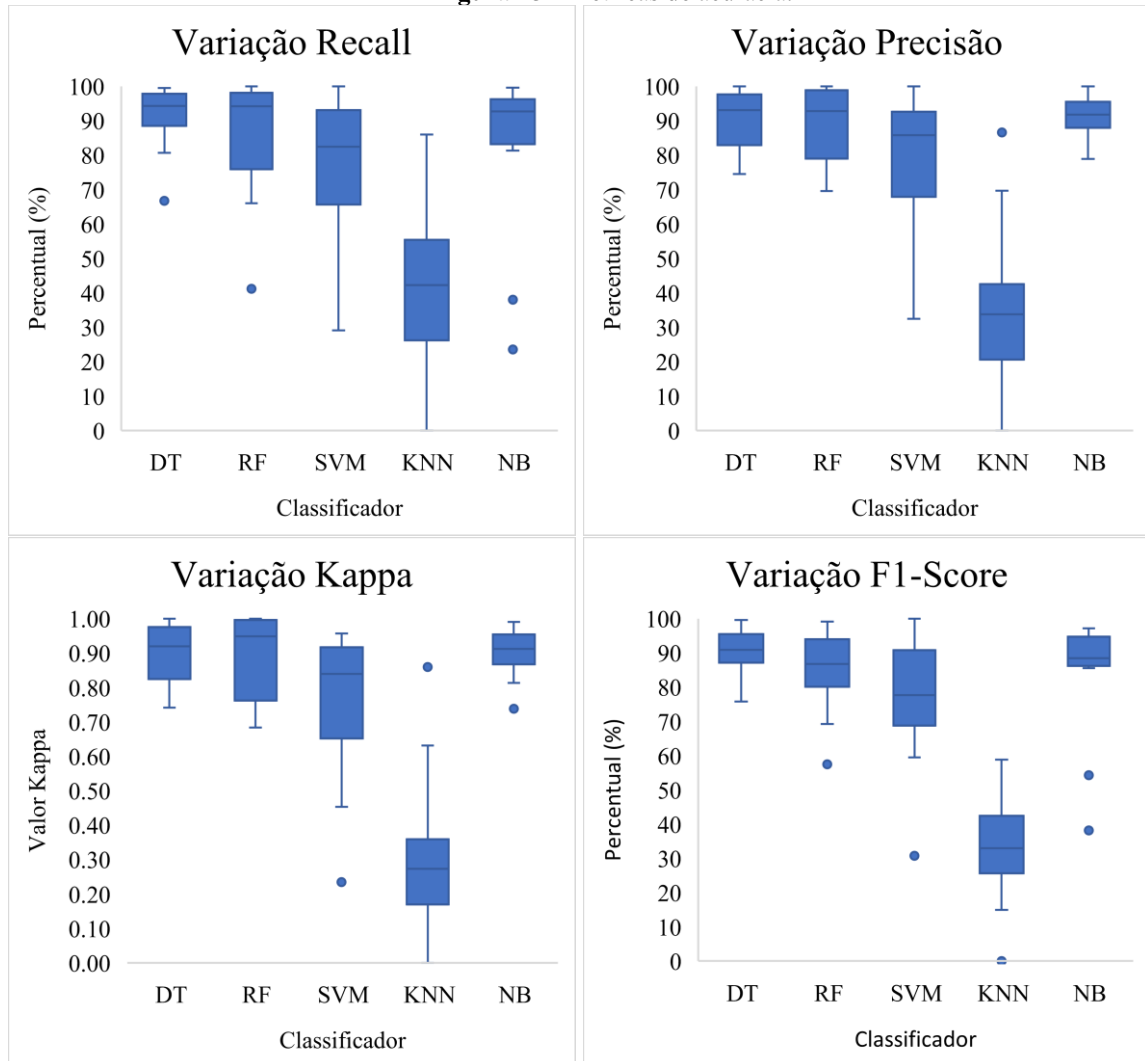
Na sequência é apresentada a discussão sobre a etapa de avaliação do desempenho dos algoritmos *Decision Tree*, *Random Forest* (RF), *Naive Bayes*, *Support Vector Machine* e *K-Nearest Neighbors*. Lembrando que essa avaliação foi realizada por meio das amostras orientadas por objeto de referência dos padrões de uso e cobertura da terra (identificados a partir de visitas a campo). Os dados da imagem da constelação de Satélite Pléiades foram processados no aprendizado de máquina (criação do modelo do algoritmo), por meio do OTB e avaliados por meio do plugin SCP, ambos no Qgis.

## 5.1 Desempenho dos modelos

Após a identificação e seleção das amostras de treinamento, o conjunto de dados foi dividido aleatoriamente com uma proporção de 70:30 para treinamento e validação, respectivamente. O número de registros no treinamento do algoritmo e o conjunto de dados de validação somaram 404 e 174 observações, respectivamente. Para os cinco algoritmos aplicados foram utilizadas as mesmas amostras de treino e validação.

Dos classificadores utilizados para a classificação de uso e ocupação do solo, o que apresentou melhor desempenho foi o modelo de classificação gerado através do algoritmo *Decision Tree*. O mesmo apresentou melhor desempenho para a classificação global e para a maioria das classificações individuais das doze subclasses avaliadas. As métricas avaliadas para cada subclasse foram: *Recall*, *Precisão*, *Kappa* e *F1-Score* (Figura 15).

**Figura 15** - Métricas de acurácia.



Fonte: Da Autora (2022).

Os resultados das métricas utilizadas para avaliar o desempenho dos classificadores estão apresentados na Figura 16. Num processo de classificação da cobertura do solo, a situação ideal é aquela em que os modelos fazem a classificação correta de cada classe. Todavia, na prática, os modelos podem classificar pixels de uma classe de interesse, mas que pertencem a outra classe (falso positivo), ou classificar o pixel de uma dada classe de interesse como sendo de outra classe (falso negativo). As métricas são uma forma de avaliar essas ocorrências,

cabendo ao usuário avaliar qual tipo de erro na classificação é aquele que pode trazer maior prejuízo ao produto desejado. Na sequência serão apresentados os resultados de desempenho de cada modelo, a partir das métricas de avaliação.

*Decision tree (DT)*: Com base no indicador RECALL, o modelo DT obteve desempenho ótimo para a quase totalidade das 12 classes. Os piores desempenhos foram para as classes Grama (66,7%), seguida de Solo exposto (80,6%) e Telhado de cerâmica clara (81,3%). Para a Precisão, os indicadores com pior desempenho foram para telhado de cerâmica clara (74,5%), seguido de *Árvore* e Calçada, ambos em torno de 78%. Para F1 score, os dois piores desempenhos estão associados às classes de cobertura Grama e Telhado de cerâmica clara, com 75,7% e 77,7%, respectivamente. Para o índice Kappa, os piores desempenhos estão associados às classes Telhado de cerâmica clara (0,74), *Árvore* (0,75) e Calçada (0,78). Para as demais classes de cobertura, a classificação pode ser ótima, com base nos valores das métricas.

*Random Forest (RF)*: O modelo apresentou desempenho no indicador recall entre 66% e 100 % sendo classificados com pontuação máxima telhado cerâmica escura e sombra e a menor pontuação foi para a classe telhado metálico (41%). No indicador precisão, o menor valor foi de 69,5 % para a classe telhado cerâmica escura e o maior valor para as classes árvores, piscina e solo exposto com o valor de 100%. Já para o índice kappa, as classes piscina, solo exposto e calçada obtiveram classificação máxima obtendo o valor 1,0, classificando como excelente.

*Naive Bayes (NB)*: O indicador recall para o modelo apresentou melhores resultados para as classes asfalto e telhado cinza, ambas 99%, e o menor valor foi para a classe água 23%. A métrica precisão variou entre 100 e 82%, sendo que a classe árvore obteve o valor de precisão de 82%, e a classe água, de 100%. No índice Kappa, o pior desempenho foi para classe solo exposto, com o valor de 0,74, seguido da classe arvore (0,81), enquanto que as demais classes obtiveram desempenho ótimo conforme os valores da métrica.

*Support Vector Machine (SVM)*: Para a métrica *recall*, os valores variaram entre 100%, valor atribuído para as classes piscina e telhado de cerâmica escura, e o menor valor de recall, de 29%, para a classe asfalto. Os valores de precisão variam de 32%, valor adquirido para classe asfalto, e 100%, para as classes piscina e telhado de cerâmica escura. O valor de Kappa foi de 0,23 para a classe asfalto, sendo uma classificação com desempenho fraco, assim como a classe solo exposto, com o valor de 0,46. As demais classes podem ser classificadas entre bom e ótimo.

*K-Nearest Neighbors (KNN)*: Este é o modelo que apresentou menor desempenho nas métricas de acurácia. Os valores de recall foram de 85% para a classe piscina e 0% para a classe água, com a média de 42% para este modelo. Os valores de precisão foram de 69% para a classe

sombra e 0% para a classe água. Os valores de Kappa ficaram entre 0 e 0,6 e, com exceção da classe sombra, todas as outras classes foram classificadas entre fraca e regular.

As métricas *Recall* e Precisão podem ser usadas para medir de que forma o modelo consegue reconhecer as instâncias positivas (*recall*) ou negativas (Precisão).

A pontuação *F1-Score* combina os valores de precisão e *recall* em relação a uma classe. O *F1-Score* pode ainda ser interpretado como uma média ponderada da Precisão e *Recall*, onde uma pontuação F1 atinge seu melhor valor em 100% e pior em 0. É uma média harmônica entre o *Recall* e Precisão, que está muito mais próxima dos menores valores do que uma média aritmética simples (RODRIGUES, 2019). Conforme mostrado anteriormente, para as subclasses avaliadas, os melhores valores dessa métrica foram para a classificação realizada através do algoritmo DT, apresentando menor dispersão e uma média de 90,8%, sendo um valor ótimo para a métrica.

Em relação ao índice Kappa, os melhores desempenhos foram para os algoritmos NB e DT, ambos apresentando uma média de 0,9. O índice *Kappa* é a medida relativa do grau de concordância entre a classificação digital realizada e a fonte de referência (CAMPBELL e WYNNE, 2011). Os valores do índice *Kappa* variam de -1 a 1, onde 0 indica sem concordância e 1 concordância perfeita. Os valores do índice *Kappa* são normalmente utilizados quando se compara distintos classificadores.

No caso deste trabalho, espera-se oferecer subsídio ao tomador de decisão quando há interesse em avaliar o potencial de instalação de SbN em uma dada área. Nesse sentido, o excesso de falso positivo pode ser um agravante, uma vez que pode induzir o gestor a elaborar projetos e empenhar recursos em áreas que podem não corresponder à expectativa de sucesso decorrente da implantação de uma SbN. Portanto, a Precisão parece ser a métrica mais indicada para a tomada de decisão no contexto deste trabalho.

## **5.2 Mapeamento da cobertura do solo para a BH**

O método mais usual para avaliar a acurácia de maneira mais detalhada é através da matriz de confusão, também conhecida como matriz de erro. A matriz de erro do presente estudo encontra-se no Apêndice I. A denominação confusão se dá pelo fato de que, além de indicar a acurácia geral, a matriz também indica a confusão que o classificador pode gerar entre classes. Os valores apresentados na matriz correspondem à quantidade de pixels que o classificador reconheceu, ou a porcentagem dos mesmos (SILVA, 2020; CAMPBELL e WYNNE, 2011).



A principal medida utilizada para expressar a acurácia global é a exatidão global, calculada a partir da matriz de erro. Tal matriz se refere ao somatório dos valores obtidos na diagonal principal divididos pela quantidade de pixels avaliados, identificando a proporção dos pixels que tiveram a classificação correta (SILVA, 2020).

A norma brasileira ET-CQDG (DSG, 2016b) adota a acurácia geral, com base na matriz de confusão, para avaliar a qualidade temática de um mapa de classificação do uso do solo.

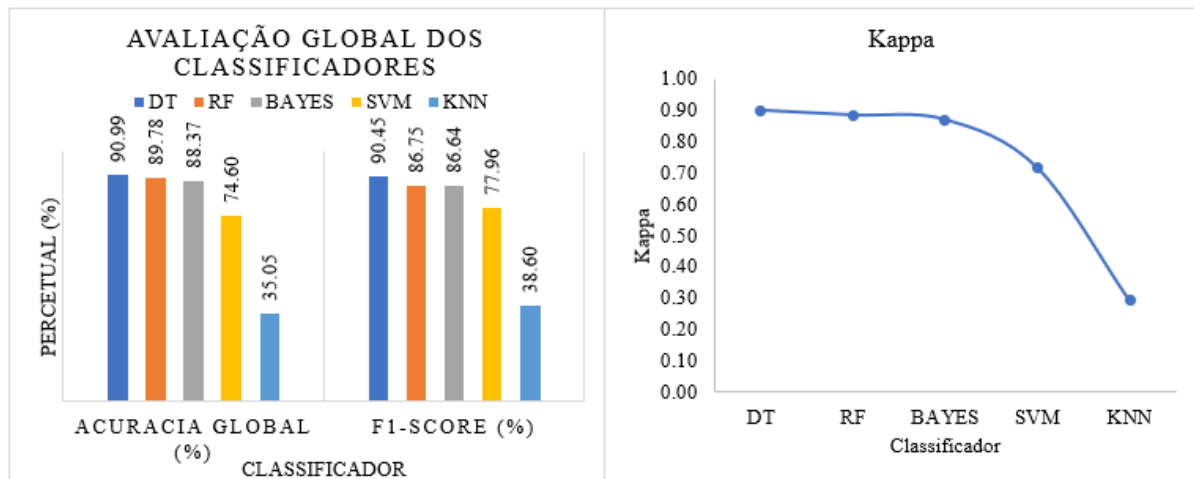
Além da avaliação da acurácia de cada subclasse, utilizando quatro métricas de avaliação (*Recall*, *Precisão*, *Kappa* e *F1-score*), foi estabelecida uma avaliação global para os classificadores, avaliando acurácia global, *F1-Score* e *Kappa*. Os resultados constam na figura 16.

Dos cinco classificadores avaliados na classificação global, o algoritmo de melhor desempenho foi o DT (*Decision Tree*). Considerando a acurácia global (G) e o índice *Kappa*, os algoritmos DT e RF tiveram os resultados mais significativos. Entretanto, individualmente o algoritmo DT apresentou um desempenho considerável como mostra a figura 16. Enquanto a *Random Forest* obteve um desempenho semelhante ao DT, o algoritmo K-NN apresentou desempenho inferior entre os testes.

Em um estudo realizado por Sampaio et. al (2019), referente à classificação do uso do solo de uma bacia hidrográfica por meio de algoritmos de *Machine Learning*, as melhores classificações de uso do solo foram obtidas com os algoritmos *Decision Tree* e K-NN, quando aplicados às seis primeiras bandas do Landsat 8.

Chamma et. al (2021), utilizando técnicas de *Machine Learning* para classificação de telhados em área urbana, obteve maior desempenho para os algoritmos *Random Forest* e *Redes neurais*, obtendo um valor de 0,94 para o índice *Kappa*. No caso deste trabalho, os valores obtidos para a classificação utilizando o algoritmo *Decision Tree*, foram de 90% de precisão para as três métricas de avaliação global (*Acurácia Global*, *F1-score* e *Kappa*). Na sequência vem os algoritmos *Random Forest* e *Bayes*, com desempenhos muito próximos. O pior resultado é observado para o algoritmo KNN. A figura16 representa os resultados do desempenho dos modelos avaliados.

**Figura 16-** Acurácia Global.

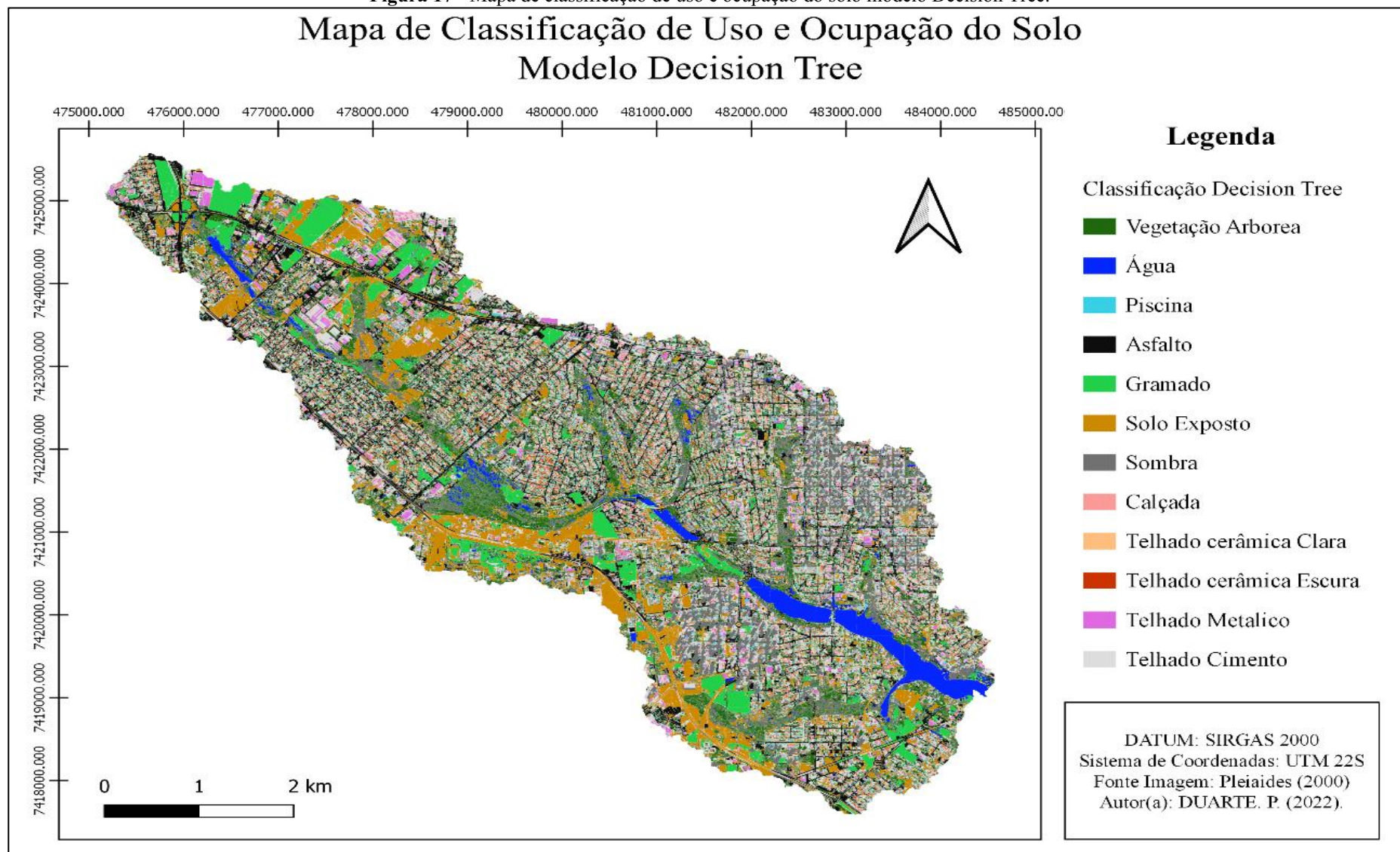


Fonte: Da Autora, (2022).

Os algoritmos de *Machine Learning* utilizados na presente pesquisa fazem parte da categoria de aprendizado supervisionado. Dentre os cinco modelos de classificação aplicados, o que apresentou maior desempenho entre os métodos de classificação foi a *Decision Tree*, seguido do *Randon Forest*. A partir desses resultados, foram estimadas as áreas para as doze subclasses consideradas para a avaliação de uso e ocupação do solo. Somadas, essas áreas equivalem aos 30 km<sup>2</sup> da parte da Sub Bacia do Ribeirão Cambé em análise.

A figura 17 representa o mapa de classificação de uso e ocupação do solo utilizando o algoritmo *Decision tree* que é a técnica que mostrou melhor desempenho no caso deste trabalho. É também uma das técnicas de sensoriamento remoto mais utilizadas em outros trabalhos de mapeamento, pois gera modelos preditivos com alta precisão e fácil interpretação dos resultados (Yang et al., 2014; Vorovencii, 2015; Lamchin et al., 2016).

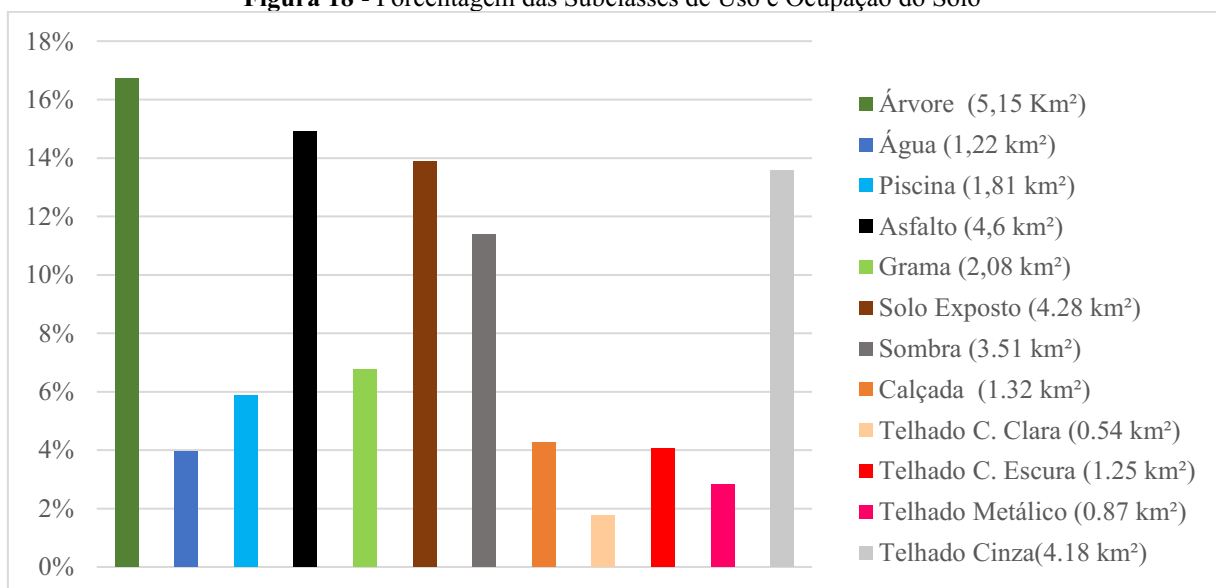
Figura 17 - Mapa de classificação de uso e ocupação do solo modelo Decision Tree.



Fonte: Da Autora, (2022).

Com os resultados obtidos foram quantificadas as participações individuais das classes utilizadas na classificação. Dos 30 km<sup>2</sup> da área de estudo, 17% (5,15 km<sup>2</sup>) desse valor foi classificado como vegetação arbórea e arbustiva. Esta fração de área também inclui remanescentes de floresta em torno dos cursos d'água. Em segundo lugar ficou o asfalto, representando 15% (4,6 km<sup>2</sup>) da área total. As áreas correspondentes à classe telhado cinza ficou com 14% (4,28 km<sup>2</sup>). A classe telhado cinza engloba telhados de amianto e áreas cimentadas da cobertura de edifícios. Solo exposto apresentou 14% da área (4,28 km<sup>2</sup>), enquanto as classe água e calçada apresentaram, cada uma, (4%) da área. A classe piscina representou 6% (1,81 km<sup>2</sup>), sendo a classe que apresentou maior confusão com as outras classes. Os alvos dessa classe são muito pequenos e isso, provavelmente, pode ter causado confusão no classificador. Para as demais classes, gramado ficou com 7% (2,08 km<sup>2</sup>), telhado claro 2% (0,54 km<sup>2</sup>), telhado escuro 4% (1,25 km<sup>2</sup>) e telhado metálico 3% (0,87 km<sup>2</sup>) (Figura 18). Cabe ressaltar ainda que uma boa parte apresenta sombreamento, o que deu à classe sombra 11% de participação.

**Figura 18 - Porcentagem das Subclasses de Uso e Ocupação do Solo**



Fonte: Da Autora, (2022).

Por meio das tecnologias avançadas de geoprocessamento, sensoriamento remoto e da aprendizagem de máquina aplicada a análise de imagem de satélite de altíssima resolução, foi possível identificar, separar e quantificar as classes de ocupação do solo presentes na sub-bacia do Ribeirão Cambé.

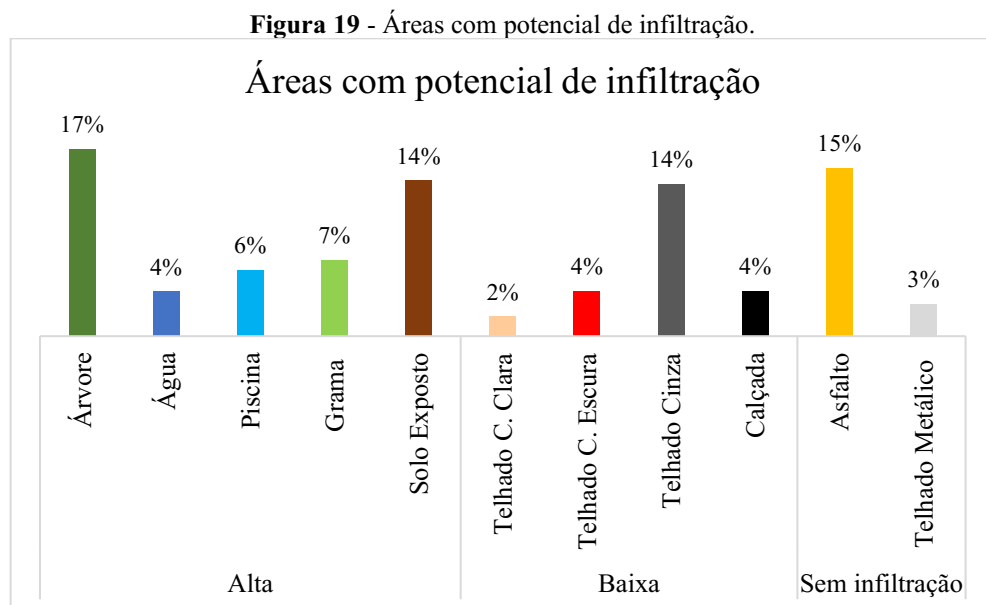
A área em estudo apresenta 22% dos seus 30 km<sup>2</sup> com elementos impermeáveis como asfalto, calçadas e telhado metálico. Esse percentual expressivo de elementos da superfície que

não absorvem água confirma se tratar de uma área muito antropizada e que vem sofrendo com problemas de alagamento, principalmente em dias de chuva intensa.

Cabe lembrar que a busca por soluções de problemas urbanos decorrentes do mau planejamento no passado vem se tornando algo recorrente pela gestão pública. Esse interesse serve tanto para mitigar as dificuldades já existentes, como também para evitar que ocorra problemas semelhantes futuramente. Nesse sentido, o uso do sensoriamento remoto, geotecnologias e técnicas de inteligência artificial são grandes aliados para uma boa gestão de planejamento, pois auxilia na tomada de decisão, otimiza o uso do tempo das visitas a campo para levantamentos e oferece produtos seguros, precisos e confiáveis.

### 5.3 Potenciais Aplicações do Estudo

A identificação das classes de ocupação do solo, por técnicas de sensoriamento remoto e de algoritmos de *Machine Learning*, auxilia na tomada de decisão em questões de adequação e planejamento urbano. Com a aplicação dessas técnicas, o presente trabalho identificou as áreas com diferentes potenciais de infiltração (Figura 19) em parte da microbacia do Ribeirão Cambé no município de Londrina. A área de estudo contém uma cobertura vegetal de 17% dos seus 30 km<sup>2</sup>. Analisando o mapa de classificação obtido, nota-se o grande corredor vegetado existente e seu enorme potencial para a implantação de SbN. Diferentemente das demais classes, as áreas vegetadas possuem potencial de aproveitamento imediato, através da instalação e manejo apropriado de espécies.



Fonte: Da autora, (2022).

Observou-se também que a área edificada, sem capacidade de infiltração ou baixa capacidade de infiltração, como asfalto e telhados metálicos, ou seja, áreas impermeáveis, representam fração considerável da área de estudo, totalizando 18% da área. As classes telhados apresentam baixa capacidade de infiltração e representam 18% da área. A classe calçada apresenta baixo percentual de infiltração, representando 4% da área e, mesmo sendo feitas de material impermeável, muitas calçadas têm espaços livres os quais proporcionam infiltração de água, como por exemplo, a área livre onde é realizado o plantio de árvore. Os telhados representam 18% da área, sendo os telhados de cinza, muitas vezes na parte superior de edifícios, com grande potencial para receber intervenções paisagísticas como, por exemplo, telhados verdes. As técnicas de SbN relacionadas à captação da água de chuva, ou que promovam o retardo em seu movimento em direção aos corpos hídricos, podem minimizar o impacto dessas áreas nos problemas relacionados à drenagem urbana. Todavia, para a tomada de decisões no planejamento urbano, torna-se necessário o desenvolvimento de estudos complementares ao presente trabalho, como, por exemplo, o regime de precipitação na região, entender a hidrologia local, além de estudos específicos de soluções baseadas na natureza que sejam resilientes, financeiramente viáveis e sustentáveis.

## 6 CONCLUSÃO

O presente estudo avaliou o uso e ocupação do solo na área de drenagem do Lago Igapó, no município de Londrina – PR. Esta avaliação foi feita a partir da combinação de imagens de satélite com elevado nível de detalhamento espacial e técnicas de Machine Learning. A classificação supervisionada orientada por objeto foi adotada na construção do mapa das classes de cobertura. Foram avaliados o desempenho dos modelos *Decision Tree*, *Random Forest (RF)*, *Naive Bayes*, *Support Vector Machine* e *K-Nearest Neighbors*, que permitiram as seguintes conclusões:

- 1) O algoritmo *Decision tree* apresentou o melhor desempenho, com valores de Kappa, Precisão, *Recall* e *F1-Score*, entre 90 % e 100 % de precisão;
- 2) *K-Nearest Neighbors* foi o algoritmo com pior desempenho. Neste caso, telhado em cor escura foi a única classe que apresentou valor acima de 80% na precisão. As demais classes apresentaram valores abaixo de 70%.
- 3) As classes identificadas por Árvores/arbustos e Grama representam quase um quarto da área de estudo, indicando um enorme potencial para implantação imediata de SbN, uma vez que essas áreas já são por natureza aptas a receber e infiltrar água da chuva;
- 4) Telhados, incluindo os metálicos, representam fração equivalente às áreas verdes (24%). Embora sejam áreas mais complexas, devido sua natureza funcional, podem oferecer grande potencial para a captação ou retenção da água de chuva, pelo menos em alguns dos tipos de telhados analisados.
- 5) A fração de solo exposto (14%) também é expressiva e chama a atenção pelo fato de serem áreas com potencial para edificação. É justamente dessas áreas, quando passam pelo processo de edificação, que vem a maioria do sedimento transportado para dentro do lago Igapó;
- 6) Asfalto e calçada, com 19% da área, são provavelmente as classes com menor potencial para implantação de SbN. Ainda assim, existe a possibilidade da adoção de pavimentos permeáveis nos calçamentos e em áreas de asfalto destinadas ao estacionamento de veículos em vias de pouco movimento.

## REFERÊNCIAS

- ADDAN, D. **Support Vector Machine**. Unibrasil, 34p. 2019.
- ALMAZROUI, M., MASHAT, A., ASSIRI, M. E., & BUTT, M. J. Application of landsat data for urban growth monitoring in Jeddah. **Earth Systems and Environment**, v.1, p. 25, 207.
- ALVARENGA JÚNIOR, Wagner José. Métodos de otimização hiperparamétrica: um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária. 2018.
- ALVES, C. A; CONCEIÇÃO, A. E. P. Levantamento do uso e ocupação do solo por meio de imagens TM-Landsat-5 e ADS-80 para o município de Manaus/AM. João Pessoa. **Anais XVII Simpósio Brasileiro de Sensoriamento Remoto**, 8p. 2015.
- AMARAL, Fernando. **Aprenda mineração de dados: teoria e prática**. Alta Books Editora, 2016.
- ANDREOLA, Rafaela. **Support Vector Machines na classificação de imagens hiperespectrais**. Porto Alegre, 2009. Monografia (Dissertação) – Programa de Pós-Graduação em Sensoriamento Remoto Universidade Federal do Rio Grande do Sul, 2009.
- APRENDIZADO DE MAQUINA. **Os Três Tipos de Aprendizado de Máquina**. Disponível em: <https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>. Consultado em: 08/02/2022.
- ARANHA RIBEIRO, A. R. S. CENTENO, S. J. Classificação do Uso do Solo Utilizando Redes Neurais e o Algoritmo MAXVER. *In*: Simpósio Brasileiro de Sensoriamento Remoto, **Anais[...]**, Foz do Iguaçu, 2001, Instituto Nacional de Pesquisas Espaciais, v. 20, p. 1341-1348, 2001.
- ARAUJO, Silva Rafael, **Microbacia do Ribeirão Cambé-Londrina-PR: Levantamento Ambiental Utilizando Técnicas de Geoprocessamento e Sensoriamento Remoto**. 2004.
- ARLOT, S; CELISSE, A. A survey of cross-validation procedures for model selection. **Statistics surveys**, v. 4, p. 40-79, 2010.
- ARONOFF, S., Geographic Information Systems: A Management Perspective. **WDL Publications**, Ottawa, Canada, 1989.
- BÁRBARA M. RIBEIRO, G. Sensoriamento Remoto para Mapeamento Urbano: Classificação de Cobertura, Uso e Ocupação do Solo. Santos. **Anais XIX Simpósio Brasileiro de sensoriamento Remoto**. 4p. 2019.
- BARNESLEY, M.J.; MOLLER-JENSEN, L. E BARR, S.L. Inferring urban land use through spatial and structural pattern recognition. Londres. **Remote Sensing and Urban Analysis**. 2001.
- BARROS, F. V. M. ARCHELA, S. R. BARROS, F. N. O. THÉRY, H. MELLO, A. N. BATISTA GRATÃO, B. H. L. **Curso e (per) curso das águas**. Consultado em: <<http://www.uel.br/revistas/atlasambiental/NATURAL/CURSODASAGUAS.htm>> Acessado em: 03, março,2021.



BISHOP, Yvonne MM; FIENBERG, Stephen E.; HOLLAND, Paul W. **Discrete multivariate analysis** Cambridge. 1975.

BLASCHKE T. & KUX H. Sensoriamento Remoto e SIG Avançados: novos sistemas sensores métodos inovadores. São Paulo: **Oficina de Textos**, 2005.

BORTOLO, A. C. The production of the public free space of lago Igapó in urban context londrinense: historical analysis of photos and data as tools for the construction of the study. Presidente Prudente, **Revista Geografia em Atos**, v.2, n. 11, p. 97-109, 2011.

BRADSKI, Gary; KAEHLER, Adrian. **Learning OpenCV: Computer vision with the OpenCV library**. " O'Reilly Media, Inc.", 2008.

BRANDT. TSO MATHER, P. **Classification Methods for Remotely Sensed Data**. ed.2, London, CRC Press, 2009, 352 p.

BREIMAN, L. Random forests. Machine learning, **Springer**, v. 45, n. 1, p. 5–32, 2001.

BURROUGH, P. A. Principles of Geographical Information Systems for Land Resources Assessment. Oxford: **Clarendon Press**, 1986.

CÂMARA, G.; FREITAS, U.M.; SOUZA, R.C.M.; GARRIDO, J. SPRING: Integrating Remote Sensing and GIS by Object-Oriented Data Modelling. **Computers and Graphics**, vol. 15, n.6, July 1996.

CÂMARA, G; DAVIS, C. MONTEIRO, V. M. A. **Introdução à Ciência da Geoinformação(2001)**. Disponível em: <http://mtc-m12.sid.inpe.br/col/sid.inpe.br/sergio/2004/04.22.07.43/doc/publicacao.pdf>. Consulto em: 28 de agosto de 2020.

CAMARGOS, A. L. SILVA, C. C. MORAES, F. M. PRUDENTE, N. C. Aplicação do Modelo Digital de Terreno (MDT) Advanced Elevation Series (AES) em parte do município de Formosa, GO. *In: XVII Simpósio Brasileiro de Sensoriamento Remoto, 2015, João Pessoa, Anais[...]*, Brasil, Paraíba: INPE, 2015, p. 4559-4566.

CAMPBELL, James B; WYNNE, Randolph H. **Introduction to Remote Sensing**. 5th ed. The Guilford Press, New York, 2011.

CARDOSO, Machado Priscila. **Análise das Características Físicas e Ambientais da Bacia do Córrego Tucanos em Londrina-PR**. 2013, Monografia (Bacharel em Geografia) – Universidade Estadual de Londrina, Londrina, 2013.

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

CHAMMA, Willian Douglas Sbitkowski et al. Aprendizado de máquina aplicado em imagens de satélite para classificação de telhados Machine learning applied to satellite imagery for rooftop classification. **Brazilian Journal of Development**, v. 7, n. 7, p. 72558-72576, 2021.

CHANG, Mein Mieko; PINHEIRO, Hemerson Donizete. Análise temporal do uso e cobertura do solo de uma bacia hidrográfica urbana do município de Londrina/PR.

CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, v. 49 n. 12, p. 1671-1678, 1991.

CORNELLI, R.; SCHNEIDE, E. V.; BORTOLINI, A. T.; CEMIN, G.; SANTOS, M. G. Análise da Influência do Uso e Ocupação do Solo na Qualidade da Água de Duas Sub-Bacias Hidrográficas do Município de Caxias do Sul. Caxias do Sul, **Scientia Cum Industria**. V.4, N. 1, 1 — 14, 2016.

COSTA, P. C. E., AUGUSTO, C. R., SEABRA, S. V. Análise da eficiência dos índices Built-up e NDBI para classificação de áreas urbanas em imagens Landsat 8 OLI. *In: XVIII Simpósio Brasileiro de Sensoriamento Remoto, 2017, Santos, Anais [...] Brasil, São Paulo, INPE, p.6632-6639.*

COULIBALY, P. J. S. H. HONORÉ, J. T. C. NAGA, C. KOUADIO, A. C. K. DIDI, M. R. S. DIEDHIOU. A SAVANE, I. Groundwater exploration using extraction of lineaments from SRTM DEM and water flows in Béré region. **The Egyptian Journal of Remote Sensing and Space Science**, 2020.

COWEN, D. J. GIS versus CAD versus DBMS: what are the differences? **Photogrammetric Engineering and Remote Sensing**, v. 54, p. 1551-1554, 1988.

CRÓSTA, A. P. **Processamento Digital de Imagens de Sensoriamento Remoto**. Campinas. UNICAMP edição 4, 2002.

DE ESPINDOLA, Giovana Mira; CÂMARA, Gilberto. Ajuste de parâmetros no algoritmo de segmentação de imagens do software eCognition®. 2007.

DEVECCHI, Alejandra Maria et al. Desenhando cidades com Soluções baseadas na Natureza. **Parcerias Estratégicas**, v. 25, n. 50, p. 217-234, 2021.

DIAS, Santos Furtado Danielle. **Representações de séries temporais baseadas em imagens para classificação de imagens de satélite**, 2020, Monografia (Tese obtenção do título de Doutor em Ciência da Computação) – Universidade Estadual de Campinas, Campinas, 2020.

DOURADO, A. A. PINHEIRO, S. L. Utilização de Técnicas de Geoprocessamento para Análise do Uso e Ocupação do Solo da Laguna do Sargento em Icaraízinho de Amontada-CE. **Revista Homem espaço e tempo**, v12. n. 1, 15p.2018.

EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Informação Tecnológica: Geotecnologias e Geoinformação: o produtor pergunta, a Embrapa responde**. Brasília, Embrapa, 2014.

FANTINEL, Roberta Aparecida et al. Procedimentos de aprendizagem de máquina para análise de padrões espaciais com o uso da plataforma KNIME. 2020.

FITZ, P. R. **Geoprocessamento sem complicação**. São Paulo: Oficina de Textos, 2008.

FRAGA, Raiza Gomes; SAYAGO, Doris Aleida Villamizar. Soluções baseadas na Natureza: uma revisão sobre o conceito. **Parcerias Estratégicas**, v. 25, n. 50, p. 67-82, 2021.

FREITAS e SANTANA, Machine Learning: Desafios Para Um Brasil Competitivo. **Computação Brasil**. v. 39, ed. 01. P. 7-9. 2019.

FUKUNAGA, Keinosuke. **Introduction to statistical pattern recognition**. Elsevier, 2013.

GALHOTE, Bilotti Luiz Victor. **Diagnóstico das condições ambientais segundo os parâmetros do IQA em um corpo hídrico lântico – Lago Igapó I**. 2019 Monografia (Bacharel em Engenharia Ambiental) Universidade Tecnológica Federal do Paraná, Londrina, 2019.

GAROFALO, T. F. D., MESSIAS, G. C., LIESENBERG, V. BOLFE, L. E. FERREIRA, C. M. Análise comparativa de classificadores digitais em imagens do Landsat-8 aplicados ao mapeamento temático. **Pesquisa Agropecuária. brasileira**, Brasília, v.50, n.7, p.593-604, 2015.

GASPARINI, C. A. K.; LYRA, B. G.; FRANCELINO, R. M.; DELGADO, C. R.; JUNIOR, O. F. J.; FACCO, A. G. Técnicas de Geoprocessamento e Sensoriamento Remoto Aplicadas na Identificação de Conflitos do Uso da Terra em Seropédica-RJ. Rio de Janeiro. **Floresta e Ambiente**. v.20, n.3, pp.296-306.2013.

GOEBEL, Michael; GRUENWALD, Le. A survey of data mining and knowledge discovery software tools. **ACM SIGKDD explorations newsletter**, v. 1, n. 1, p. 20-33, 1999.

GOMES, A. A. M. **O que é e para que serve o Geoprocessamento**. Disponível em: <https://www.unifai.edu.br/pesquisa/publicacoes/professores/sequenciais/o-que-e-e-para-que-serve-o>. Consultado em: 05 de out de 2020.

GONÇALVES, André Ricardo. Máquina de vetores suporte. **Acesso em**, v. 21, 2010.

GONG P, WANG J, YU L, et al. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. **Int J Remote Sens**, v.34, p.54, 2013.

GUEDES, Ademir Rafael Marques; GUIMARÃES, Victor Luiz. Sistema de Reconhecimento Baseado em Random Forest para Caracteres de Captchas. **Departamento de Computação, Universidade Federal de Ouro Preto, Ouro Preto, MG**, v. 5, 2021.

HADJIMITSIS, D. G., PAPADAVID, G., AGAPIOU, A., THEMISTOCLEOUS, K., HADJIMITSIS, M. G., RETALIS, A., et al. Atmospheric correction for satellite remotely sensed data intended for agricultural applications: impact on vegetation indices. **Natural Hazards and EarthSystem Sciences**, v.10, p.89–95, 2010.

HEPNER, Georgef et al. Artificial neural network classification using a minimal training set- Comparison to conventional supervised classification. **Photogrammetric Engineering and Remote Sensing**, v. 56, n. 4, p. 469-473, 1990.

HERZOG, Cecilia Polacow. Soluções baseadas na Natureza para um novo paradigma no tratamento de esgoto em áreas urbanizadas. **Parcerias Estratégicas**, v. 25, n. 50, p. 133-158, 2021.

HOMEM, William Ludovico; UFES, PET Engenharia Mecânica. Apostila de Machine Learning.

HUANG, B. ZHAO, B. SONG, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. **Rem. Sens. Environ.**, v.214, p. 73-86,2018.

HUANG, X., WEN, D., LI, J. QIN, R. Multi-level monitoring of subtle urban changes for the megacities of China using high-resolution multi-view satellite imagery. **Remote Sensing of Environment.** v.196, p. 56-75, 2017.

HUANG, X., WEN, D., LI, J. QIN, R. Multi-level monitoring of subtle urban changes for the megacities of China using high-resolution multi-view satellite imagery. **Remote Sensing of Environment.** v.196, p. 56-75, 2017.

IBAÑEZ, Menecucci Marilyn. **Uso de redes neurais nebulosas e florestas aleatórias na classificação de imagens em um projeto de ciência cidadã.** São José dos Campos, Monografia (Dissertação) - Instituto Nacional de Pesquisas Espaciais, 2016.

IBAÑEZ, Menecucci Marilyn. **Uso de redes neurais nebulosas e florestas aleatórias na classificação de imagens em um projeto de ciência cidadã.** São José dos Campos, Monografia (Dissertação) - Instituto Nacional de Pesquisas Espaciais, 2016.

INSTITUTO AGRONÔMICO DO PARANÁ. **Atlas Climático.** Disponível em: <http://www.iapar.br/modules/conteudo/conteudo.php?conteudo=2533>. Consultado em: 15 de julho de 2020.

INSTITUTO DE PESQUISA E PLANEJAMENTO URBANO DE LONDRINA. **Hidrografia.** Consultado em: <<http://www1.londrina.pr.gov.br/dados/images>> Acessado em: 08, abril, 2021.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **LANDSAT.** Disponível em: <http://www.dgi.inpe.br/documentacao/satelites/landsat> Consultado em: 01 de setembro de 2020.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS. **MANUAIS TUTORIAIS DE GEOPROCESSAMENTO: SENSORIAMENTO REMOTO.** Disponível em: [http://www.dpi.inpe.br/spring/portugues/tutorial/introducao\\_sen.ht](http://www.dpi.inpe.br/spring/portugues/tutorial/introducao_sen.ht) Machine Learning: 19 de junho de 2022.

JAPAN AEROSPACE EXPLORATION AGENCY. **Precise Global Digital 3D Map ALOS World 3D.** Consultado em: [https://www.eorc.jaxa.jp/ALOS/en/aw3d/index\\_e.ht](https://www.eorc.jaxa.jp/ALOS/en/aw3d/index_e.ht). Acessado em: 29, março, 2021.

JENSEN, J.R. E COWEN, D.C. Remote sensing of urban/suburban infrastructure and socioeconomic attributes. **Photogrammetric Engineering & Remote Sensing**, v. 65, n. 5, pp. 611-622, 1999.

JUNIOR, A. P. A; NETO, P. C. O. O Mapeamento Da Área De Risco À Jusante Do Lago Igapó Em Londrina – Pr - Brasil, no Caso De Rompimento Hipotético Da Barragem. **Territorium**, Coimbra-PT. V 27. Nº2, pg 29-40, 2020.

KAVZOGLU, Taskin; COLKESEN, Ismail. A kernel functions analysis for support vector machines for land cover classification. **International Journal of Applied Earth Observation and Geoinformation**, v. 11, n. 5, p. 352-359, 2009.

BARSI, Á. et al. Accuracy Dimensions in Remote Sensing. **International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences**, v. 42, n. 3, 2018.

LAMCHIN, Munkhnasan et al. Assessment of land cover change and desertification using remote sensing technology in a local region of Mongolia. **Advances in Space Research**, v. 57, n. 1, p. 64-77, 2016.

LAROCCA, A. G., CARDOSO, C., DE ANGELIS, B. L. D., & DE ANGELIS, B. L. D. O Impacto da Ocupação de Fundo de Vales em Áreas Urbanas—Estudo de caso Lago Igapó Londrina—PR. **Revista Nacional de Gerenciamento de Cidades**, 5(30), 52-54. (2017).

LAWRENCE, Rick et al. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. **Remote sensing of environment**, v. 90, n. 3, p. 331-336, 2004.

LI, X.; GONG, P.; LIANG, L. A 30-year (1984–2013) record of annual urban dynamics of Beijing city derived from Landsat data. **Remote Sens. Environ**, v.166, p.78–90, 2015.

LILLESAND, T. M. KIEFER. R. W. Remot Sensing And Image Interpretatin. Crawfordsville: **J. Wiley And Sons**. 3ed. 750p. 1994.

LIU, X.; HU, G.; CHEN, Y.; LI, X.; XU, X.; LI, S.; PEI, F.; WANG, S. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. **Remote Sens. Environ**, v.209, p.227–239, 2018.

LORENZO, P. M. **Caracterização dos impactos ambientais negativos e medidas mitigatórias do processo de assoreamento do Lago Igapó, Londrina – PR**. 2011, Monografia (graduação em Tecnologia em Gestão Ambiental) - Universitário Filadélfia, Londrina, 2011.

MACHADO, M. F. L., MACHADO, P. S. C. R., ARAGÃO, R. Relação entre área impermeável e densidade demográfica com a aplicação de geoprocessamento de dados. Belém, **Congresso Técnico Científico da Engenharia e da Agronomia**, 5p., 2017.

MACHADO, Walquíria; COLAVITE, Ana Paula; STIPP, Nilza Aparecida Freres. Avaliação De Impactos Ambientais Através De Matriz De Interação Na Gleba Palhano—Londrina-Pr. Monografia (Bacharel em Geografia) – Universidade Estadual de Londrina, Londrina, 2004.

MANUAIS TUTORIAL DE GEOPROCESSAMENTO. **Instituto Nacional de Pesquisas Espaciais**. Disponível em: [http://www.dpi.inpe.br/spring/portugues/tutorial/introducao\\_sen.htMachine](http://www.dpi.inpe.br/spring/portugues/tutorial/introducao_sen.htMachine) Learning Consultado em: 01 de out de 2020.

MATHER, Paul M.; KOCH, Magaly. **Computer processing of remotely-sensed images: an introduction**. John Wiley & Sons, 2011.

MAXWELL, Aaron E.; WARNER, Timothy A.; FANG, Fang. Implementation of machine-learning classification in remote sensing: An applied review. **International Journal of Remote Sensing**, v. 39, n. 9, p. 2784-2817, 2018.

MEDRI, E. M. BIANCHINI, E. SHIBATTA. A. O. PIMENTA, A. J. **A Bacia do Tibagi**, Universidade Estadual de Londrina, ed.-, p.601, 2002.

MENDONÇA, J. C. Aplicações Ambientais Brasileiras Com Geoprocessamento E Sensoriamento Remoto. Campina Grande. **EDUFCG**, 217p.2013.

MENESES, Paulo Roberto; ALMEIDA, T. de. Introdução ao processamento de imagens de sensoriamento remoto. **Universidade de Brasília, Brasília**, 2012.

MENESES, R. P; ALMEIDA, T. Introdução ao Processamento de Imagens de Sensoriamento Remoto. Brasília. **Conselho Nacional de Desenvolvimento Científico e Tecnológico**. 276p. 2012.

MOBLEY, C. D. **Remotely sensed data characterization, classification, and accuracies**. Boca Raton: v. 1, 2016.

MONARD, Maria Carolina; BARANAUSKAS, José Augusto. Conceitos sobre aprendizado de máquina. **Sistemas inteligentes-Fundamentos e aplicações**, v. 1, n. 1, p. 32, 2003.

MOREIRA M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicações**. Viçosa. Editora UFV edição 4. 2011.

NANDAIA, MORNA. Interpretação Geográfica da Classificação do Uso e Ocupação do Solo resultante do Processamento e análise de imagens de Satélite. 2020.

NASA EARTH OBSERVING SYSTEM DATA AND INFORMATION SYSTEM. **Alos Palsar**. Disponível em: <https://asf.alaska.edu/data-sets/sar-data-sets/alos-palsar/alos-palsar-about/> Consultado em: 01 de setembro de 2020.

NASA EARTH OBSERVING SYSTEM DATA AND INFORMATION SYSTEM. **Alos Palsar**. Disponível em: <https://search.asf.alaska.edu>. Consultado em: 01 de setembro de 2020.

NEELMEIJER, J. MOTAGH, M. BOOKHAGEN, B. High-resolution digital elevation models from single-pass TanDEM-X interferometry over mountainous regions: A case study of Inylchek Glacier, Central Asia. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p.108-121,2017.

NESSHÖVER, Carsten et al. The science, policy and practice of nature-based solutions: An interdisciplinary perspective. **Science of the total environment**, v. 579, p. 1215-1227, 2017.

NETA, A. R. S., BIAS, S. E., SANTOS, M. A. C. Object-Oriented Classification and Data Mining: Mapping Urban Area with Open Source Tools. **Revista Brasileira de Cartografia**, v. 70, n.03, p.932-966, 2018.

NG, Ritchie. **F1SCORE**. Disponível em: <https://www.ritchieng.com/machinelearning-f1-score/>. Acesso em: 08 mar. 2022.

NOVO, E. M. de M. **Sensoriamento remoto: princípios e aplicações**. São Paulo. Edgard Blucher. 1989. 308p.

NOVO, E.M.L.M. Sensoriamento Remoto: Princípios e Aplicações”, 3. ed. **Edgard Blücher**, São Paulo, 2007. 363 p. 2007.

NOVO, M. L.M.E; PONZONI, J. F. INTRODUÇÃO AO SENSORIAMENTO REMOTO. **Instituto Nacional de Pesquisas Espaciais**. São José dos Campos 81p. 2001.

PANAGIOTAKIS, E., CHRYSOULAKIS, N., CHARALAMPOPOULOU, V., POURSANIDIS, D. Validation of Pléiades Tri-Stereo DSM in urban areas. **ISPRS International Journal of Geo-Information**, v.7, n.3, p.118, 2018.

PREFEITURA DE LONDRINA. **Londrina em Dados**. Disponível em: <https://www.londrina.pr.gov.br/londrina-em-dados-2018-ano-base-2017>.

QUARTAROLI, Carlos Fernando; VICENTE, Luiz Eduardo; DE ARAÚJO, Luciana Spinelli. Sensoriamento remoto. **Embrapa Monitoramento por Satélite-Capítulo em livro técnico-científico (ALICE)**, 2014.

RAHMANA, A. ABDULLAHA, M. H., TOUSIF TANZIR, T., HOSSAIN. J., KHAN, M. B., MIAH, G., ISLAM, I. Performance of different machine learning algorithms on satellite image classification in rural and urban setup. **Remote Sensing Applications: Society and Environment**. v.20, 10p., 2020.

REDDY, O. G. P. KUMAR, N. SAHU, N. SINGH, K. S. Evaluation of automatic drainage extraction thresholds using ASTER GDEM and Cartosat-1 DEM: A case study from basaltic terrain of Central India. **The Egyptian Journal of Remote Sensing and Space Science**, v. 21, 1, p. 95-104,2018.

REN, Y. LÜ, Y. COMBER, A. FU, B. HARRIS, P. WU, L. Spatially explicit simulation of land use/land cover changes: current coverage and future prospects. **Earth-Science Reviews**. v.190, p. 398-415, 2019.

RIBEIRO, B.M.G. Identificação de ocupações irregulares em áreas urbanas utilizando imagens WorldView-2 e o classificador C4.5. Bogota. **Anais XVI Simpósio Internacional SELPER**. 21 p. 2014.

RIEGEL, P.R. ALVES, D.D. BIRLEM, E. L. ROQUE· C. D; OLIVEIRA· G.G; HAETINGER, C; OSÓRIO, M.M.D; RODRIGUES, S. A. M; QUEVEDO, M. D. Classificação de Uso e Ocupação do Terra com Ênfase em Áreas Urbanas. Rio de Janeiro. **Anuário do Instituto de Geociências – UFRJ**, v.42 n.3, p377-386. 2019.

RODRIGUES, Maria Santiellas Costa et al. Oportunidades e desafios para a utilização de soluções baseadas na natureza no combate de inundações urbanas na cidade de São Paulo/SP-Brasil. 2021.

RODRIGUES, Vitor. **Métricas de Avaliação: acurácia, precisão, recall, quais as diferenças?** Disponível em: <https://vitorborbarodrigues.medium.com/>. Acesso em: 08 mar. 2022.

ROSA, R. Geotecnologias na Geografia Aplicada. Uberlândia. **Revista do Departamento de Geografia** p. 81, 2005.

RUDKE, Anderson Paulo et al. **Dinâmica da cobertura do solo para a Bacia Hidrográfica do Alto Rio Paraná**. 2018. Dissertação de Mestrado. Universidade Tecnológica Federal do Paraná.

RUIZ, C. F. L., GUASSELLI, A. L., CATEN, T. A. Object based analysis and decision tree for the classification of submetrical spatial resolution images captured by UAV. **Bol. Ciênc. Geod**, Curitiba, v. 23, n2, p.252 - 267, 2017.

SAMPAIO, Rafael João. Classificação Do Uso Do Solo De Uma Bacia Hidrográfica Por Meio De Algoritmos De Machine Learning. **Revista Engenharia, Meio Ambiente e Inovação**, Rio de Janeiro, v. 3, n. 2, p. 29-41. 2019.

SANTOS, E. PEREIRA, R. G. EMERICH, L, D, P, S. Levantamento de causas do assoreamento de um ponto do Lago Igapó 2 no município de Londrina – PR. Campo Mourão, **GEOMAE**, v.8 n. esp., p. 242 – 250, 2017.

SANTOS, R. A. Apostila de sensoriamento remoto. Alegre. **Centro de Ciências Agrárias**. Departamento de Engenharia |Rurais, Universidade Federal do Espírito Santo. 87p. 2013.

SANTOS, R.F.dos. Planejamento ambiental: teoria e prática. São Paulo: **Oficina de Textos**, 184p. 2004.

SCHORN, Victor da Silva; YAMAMOTO, Lucas Eiji. Uso do Programa Quantum Geographical Information System (QGIS) na Análise de Alvos no Sistema ASTROS. 2021.

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

SIGKDD Explor. Newsl., **Association for Computing Machinery**, New York, NY, USA, v. 1, n. 1, p. 20–33, jun. 1999.

SILVA, Mariana de Sá Rodrigues da et al. O uso de dados de sensoriamento remoto, com ênfase à tecnologia RPA, para análise e controle da ocupação urbana. 2020.

SILVA, P. J. ZAIA. E. J. Application of data geoprocessing tools in ecological studies. **International Scientific Journal**, v.14, n. 2, p. 31-41, 2019.

SMITH, T.; MENON, S.; STAR, J.; ESTES, J. Requirements and principles for the implementation and construction of large-scale geographic information systems. **International Journal of Geographical Information Systems**, v.1, n.1 p13-31, 1987.

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

SOUZA, B. R. Sensoriamento Remoto: conceitos fundamentais e plataformas. Santa Maria. **IV CEOS WGEdu Workshop**. 76p. 2010.

SOUZA, J. C. Metodologia para avaliação ambiental em bacias hidrográficas, utilizando técnicas de geoprocessamento e sensoriamento remoto. **Tese Doutorado**. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Ciência e Tecnologia Câmpus de Sorocaba. P.96. 2017.

STEFFEN, C. A; LORENZETTI, J. A; STECH, J. L; SOUZA, R. C. M. Sensoriamento Remoto: Princípios físicos, sensores e produtos e sistema Landsat. São José dos Campos, INPE, 1981.

TRABAQUINI, K. MIGLIORANZA, E. FRANÇA, V. OSVALDO C. NETO, C. P. O Uso da geotecnologia para caracterizar os cafezais no município de Londrina-PR, em relação à altimetria, declividade e tipo de solo. **Engenharia Agrícola**, v. 30, n. 6, p. 1136-1147, 2010.



UNNIKRISHNAN, A., SOWMYAV., SOMANK P. Deep AlexNet with Reduced Number of Trainable Parameters for Satellite Image Classification. **Procedia Computer Science**, v. 143, p. 931-938, 2018.

USAMA, Muhammad et al. Unsupervised machine learning for networking: Techniques, applications and research challenges. **IEEE access**, v. 7, p. 65579-65615, 2019.

VACARIO, Ester Paula Leite; MACHADO, Ginei. Transbordamentos das águas superficiais em Londrina: o caso do Ribeirão Cambé. **Os Desafios da Geografia Física na Fronteira do Conhecimento**, v. 1, p. 512-521, 2017.

VOROVENCII, Iosif. Assessing and monitoring the risk of desertification in Dobrogea, Romania, using Landsat data and decision tree classifier. **Environmental monitoring and assessment**, v. 187, n. 4, p. 1-17, 2015.

WU, H. TANG. W. BING LUO, B. ZHONGLIANG, L. V. Weather services products generation system based on GIS geoprocessing. **Computers & Geosciences**.v. 51, p.16-21, 2013.

WU, Xindong et al. Top 10 algorithms in data mining. **Knowledge and information systems**, v. 14, n. 1, p. 1-37, 2008.

YANG, Xuemei et al. Regional-scale grassland classification using moderate-resolution imaging spectrometer datasets based on multistep unsupervised classification and indices suitability analysis. **Journal of Applied Remote Sensing**, v. 8, n. 1, p. 083548, 2014.

YEPES GUARNIZO, Jhonnatan Alexander et al. Métodos supervisionados de machine learning aplicados à produtividade agrícola de cana-de-açúcar. 2021.

YUAN F., SAWAYA K.E., LOEFELHOLZ B. C., BAUER M. E. Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multi-temporal Landsat remote sensing. **Remote Sens Environ**, v.98, p.317–328,2005.

ZAIDAN, T. R. Geoprocessamento Conceitos e Definições. Juiz de Fora. **Revista de Geografia**, v.7 n.2. 7p. 2017.

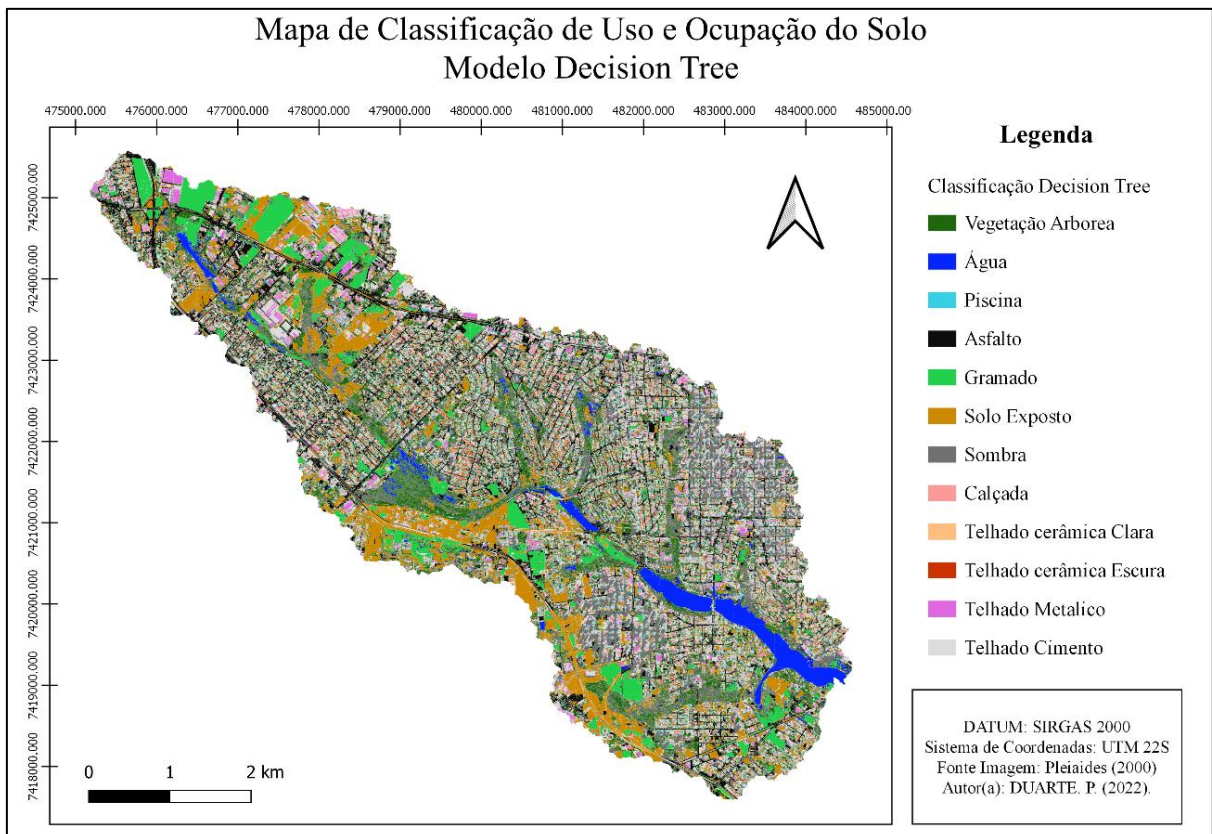
ZHANG, Harry. The optimality of naive Bayes. **Aa**, v. 1, n. 2, p. 3, 2004.

ZHAO, M. LIU, X. Reprint of: Regional risk assessment for urban major hazards based on GIS geoprocessing to improve public safety. **Safety Science**. v.97, p.112-119, 2017.

## APÊNDICE I – DECISION TREE

Matriz de confusão – *Decision Tree*

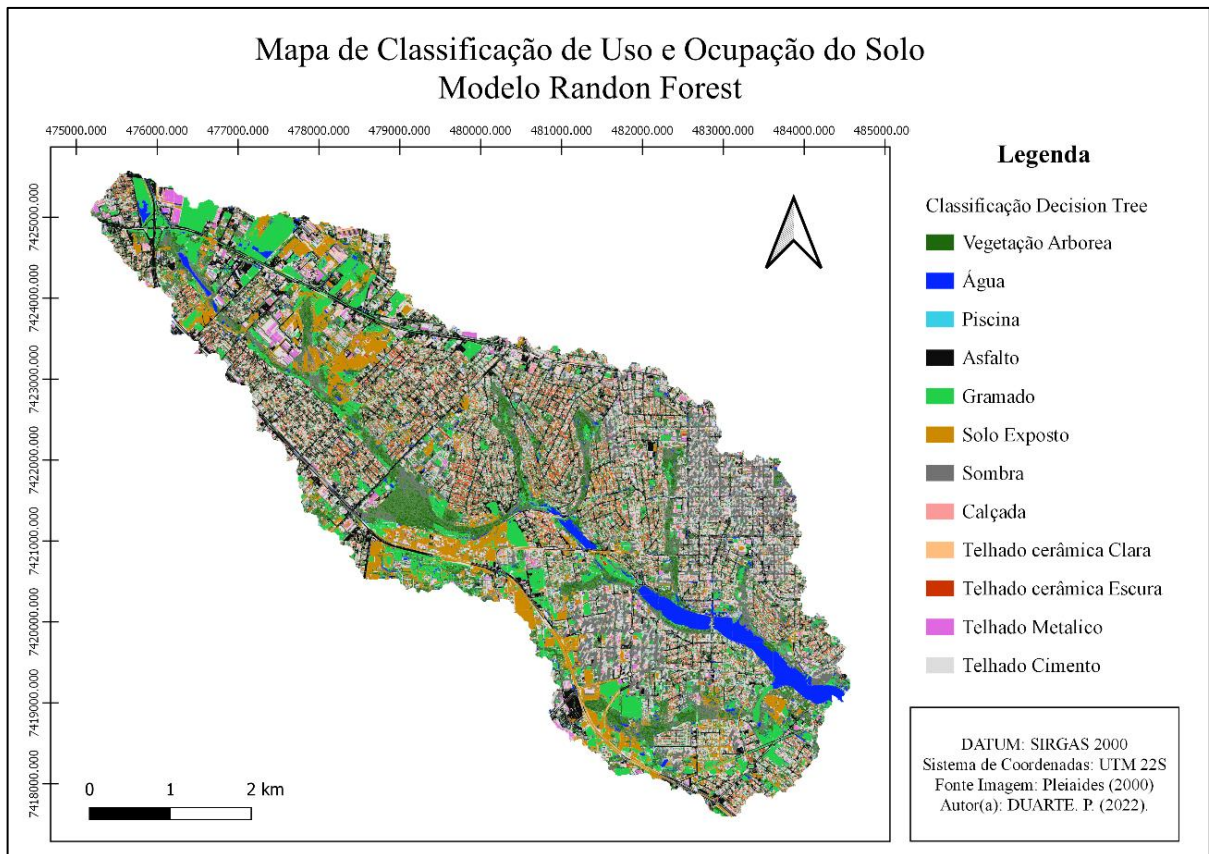
Classe	1	2	3	4	5	6	7	8	9	10	11	12	Total
1	111646	0	0	0	21019	9394	759	0	0	0	0	0	142818
2	6960	490480	0	0	0	5297	0	0	0	0	0	0	502737
3	0	0	1868	0	0	0	0	41	0	0	0	0	1909
4	0	0	0	393416	0	32919	0	0	0	0	0	0	426335
5	0	0	0	28773	657779	63692	0	0	0	0	0	0	750244
6	0	1148	0	4076	12879	504426	0	0	0	0	0	193	522722
7	260	0	0	0	0	0	359438	0	0	0	0	0	359698
8	0	0	0	0	0	0	0	1783	0	86	24	381	2274
9	0	0	0	0	0	0	0	92	1903	245	250	64	2554
10	0	0	0	0	0	0	0	0	0	2228	0	0	2228
11	0	0	0	0	0	0	0	0	1259	312	10043	280	11894
12	0	0	165	0	82	180	0	0	0	0	0	6434	6861
<b>Total</b>	118866	491628	2033	426265	691759	615908	360197	1916	3162	2871	10317	7352	2732274



## APÊNDICE II – RANDOM FOREST

Matriz de confusão – *Randon Forest*

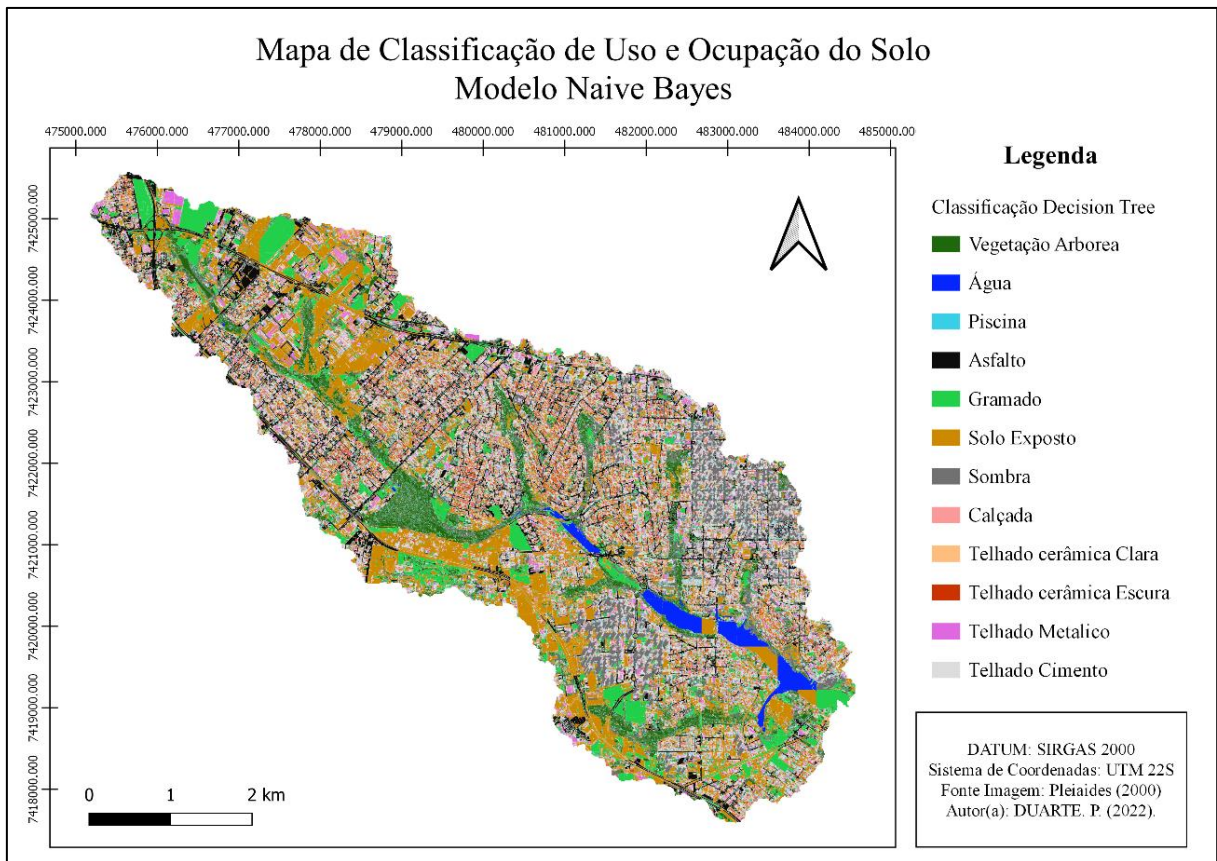
ERROR MATRIX (pixel count)													
Reference													
Classe	1	2	3	4	5	6	7	8	9	10	11	12	Total
1	86345	0	0	0	62	278	0	0	0	0	0	0	86685
2	0	466153	0	0	21019	0	4411	0	0	0	0	0	491583
3	0	0	1830	0	0	0	0	0	0	0	0	0	1830
4	0	0	0	394200	0	40202	0	0	0	0	0	0	434402
5	27292	25475	0	32065	670596	110177	0	0	0	0	0	0	865605
6	0	0	0	0	0	463994	0	0	0	0	0	0	463994
7	5229	0	0	0	0	0	355786	0	0	0	0	0	361015
8	0	0	38	0	0	0	0	1251	111	0	71	57	1528
9	0	0	0	0	0	0	0	93	2541	0	848	91	3573
10	0	0	0	0	0	1077	0	0	84	2871	96	0	4128
11	0	0	0	0	0	0	0	0	426	0	8444	0	8870
12	0	0	165	0	82	180	0	572	0	0	858	7204	9061
<b>Total</b>	118866	491628	2033	426265	691759	615908	360197	1916	3162	2871	10317	7352	2732274



### APÊNDICE III – NAIVA BAYES

Matriz de confusão – Naiva Bayes

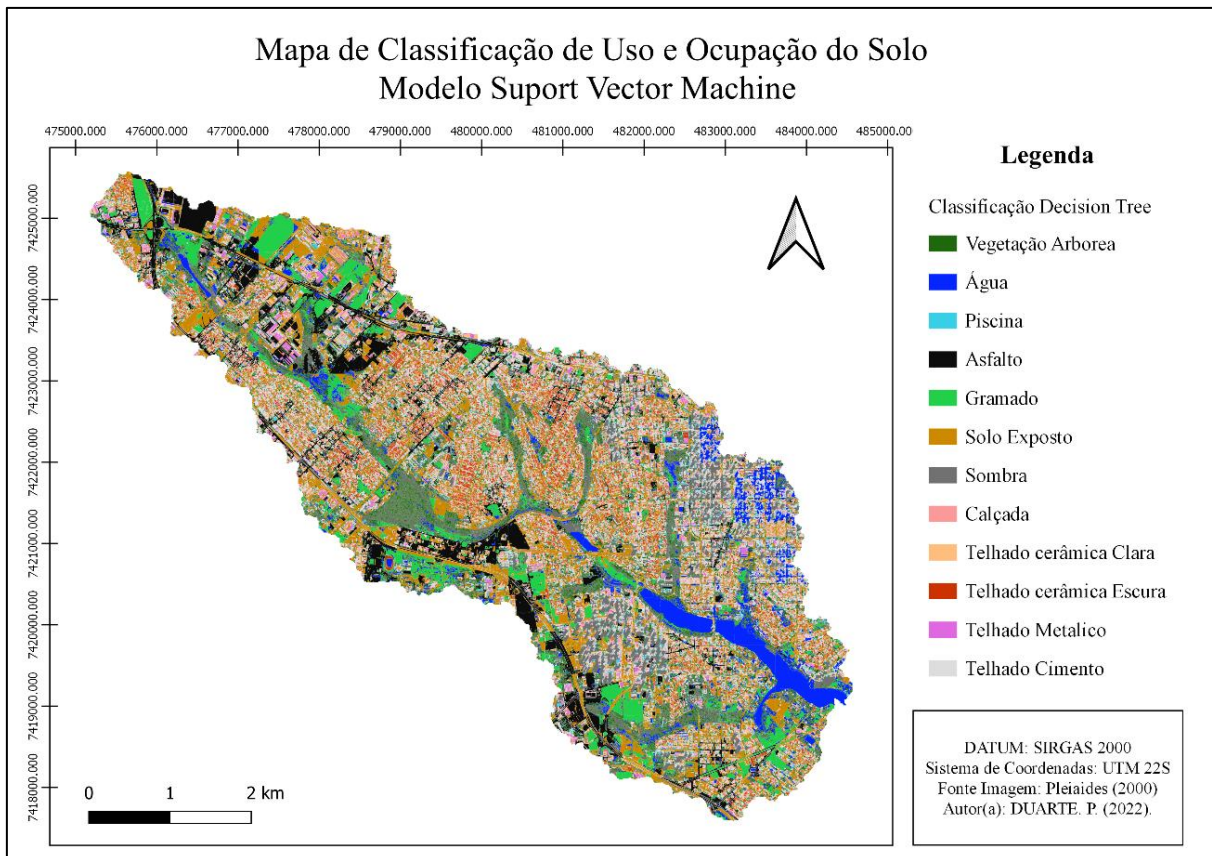
ERROR MATRIX (pixel count)															
Reference															
Classe	1	2	3	4	5	6	7	8	9	10	11	12	Total		
1	180108	36299	0	0	0	0	0	0	0	0	0	0	695	217102	
2	0	491628	0	0	0	0	0	0	0	0	0	0	0	491628	
3	0	0	2498	0	39	0	0	39	0	0	0	0	0	2576	
4	0	0	0	550451	0	65483	0	0	0	0	0	0	0	615934	
5	909	75922	0	0	903922	0	0	0	0	0	0	0	0	980753	
6	0	116953	444	1984	61049	746239	15525	0	0	61	296	3425	945976		
7	6887	0	0	0	0	0	753701	0	0	0	0	0	760588		
8	0	0	119	0	71	0	0	2615	46	198	72	0	3121		
9	0	0	0	0	0	0	0	0	3577	0	191	0	3768		
10	0	0	0	0	0	0	0	158	165	3730	38	0	4091		
11	194	0	0	0	0	0	0	67	919	0	12475	0	13655		
12	125	0	0	0	0	0	0	137	112	0	365	10514	11253		
<b>Total</b>	188223	720802	3061	552435	965081	811722	769226	3016	4819	3989	13437	14634	4050445		



## APÊNDICE IV – SUPORT VECTOR MACHINE (SVM)

### Matriz de confusão – SVM

ERROR MATRIX (pixel count)													
Reference													
Classe	1	2	3	4	5	6	7	8	9	10	11	12	Total
1	10875	1034	0	0	0	0	759	0	0	0	0	0	12668
2	3883	459258	0	0	0	0	73127	0	0	0	0	0	536268
3	0	0	2033	0	0	0	0	0	0	0	0	0	2033
4	0	0	0	258235	193555	343622	0	0	0	0	0	0	795412
5	27292	0	0	0	487675	63692	0	0	0	0	0	0	578659
6	0	0	0	168030	10529	208594	0	0	0	0	0	0	387153
7	76816	31336	0	0	0	0	286311	0	0	0	0	0	394463
8	0	0	0	0	0	0	0	1668	0	0	24	213	1905
9	0	0	0	0	0	0	0	0	2840	0	2558	170	5568
10	0	0	0	0	0	0	0	0	0	2871	0	0	2871
11	0	0	0	0	0	0	0	0	322	0	7340	0	7662
12	0	0	0	0	0	0	0	248	0	0	395	6969	7612
<b>Total</b>	118866	491628	2033	426265	691759	615908	360197	1916	3162	2871	10317	7352	2732274



## APÊNDICE V – KNN

### Matriz de Confusão

ERROR MATRIX (pixel count)													
Reference													
Classe	1	2	3	4	5	6	7	8	9	10	11	12	Total
1	15397	6481	28	0	101	0	16530	0	0	0	24	106	38667
2	0	0	0	0	0	0	0	0	0	0	0	0	0
3	6235	2722	1997	0	0	0	2021	45	0	0	0	332	13352
4	0	0	0	250527	19129	208765	10069	0	0	0	6084	0	494574
5	144737	583282	0	134688	559895	217449	615061	0	0	0	0	0	2255112
6	0	115650	0	163845	373196	367482	0	0	0	0	501	0	1020674
7	20703	11519	324	1984	12607	5785	125545	0	0	0	0	1883	180350
8	0	0	0	0	153	278	0	1322	502	812	356	779	4202
9	0	0	0	0	0	842	0	1491	2382	334	1992	3921	10962
10	0	0	0	0	0	0	0	158	257	2678	0	0	3093
11	0	0	444	1391	0	9778	0	0	1116	0	3457	4035	20221
12	1151	1148	268	0	0	1343	0	0	562	165	1023	3578	9238
<b>Total</b>	188223	720802	3061	552435	965081	811722	769226	3016	4819	3989	13437	14634	4050445

