



THE DEVELOPER'S CONFERENCE

Trilha – Data Science

Alan Delgado de Oliveira
Cientista de Dados - Poatek

Matheus Gonzaga
Cientista de Dados - Poatek



THE DEVELOPER'S CONFERENCE

Data Science para Classificação de Floresta em Imagens de Satélite

Agenda



THE
DEVELOPER'S
CONFERENCE

- Definição do Problema
- Dados de Entrada do Problema
- Anotação das Classes de Interesse
- Pré-processamento dos Dados
- Avaliação do Modelo de Classificação
- Resultado e Conclusão

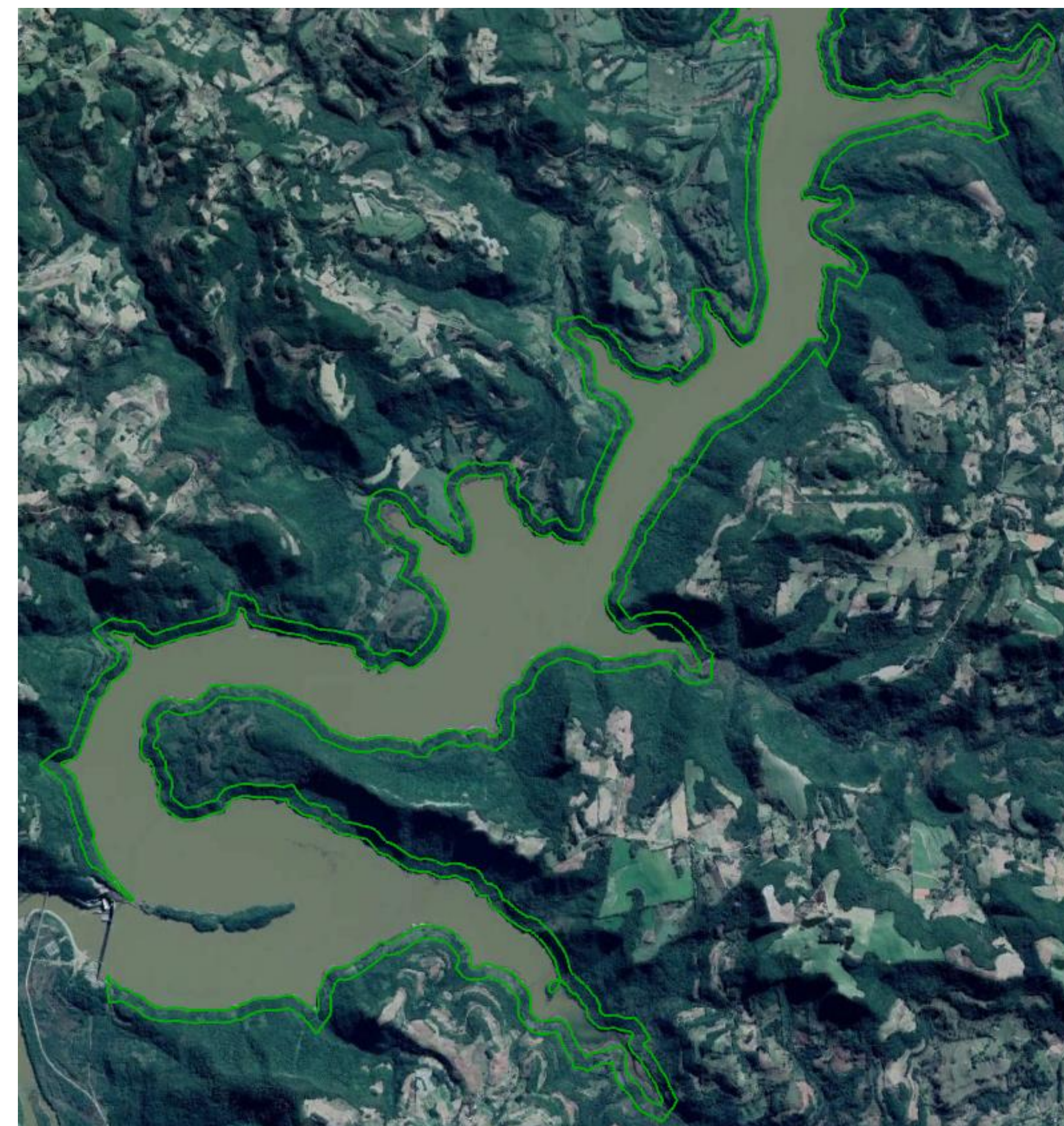
Descrição



THE
DEVELOPER'S
CONFERENCE

Realizar o mapeamento e classificação da vegetação (segmentação semântica) na Área de Preservação do reservatório da UHE Dona Francisca, considerando os seguintes estágios sucessionais:

- Vegetação campestre;
- Floresta em estágio inicial;
- Floresta em estágio médio e avançado;
- Sombra.

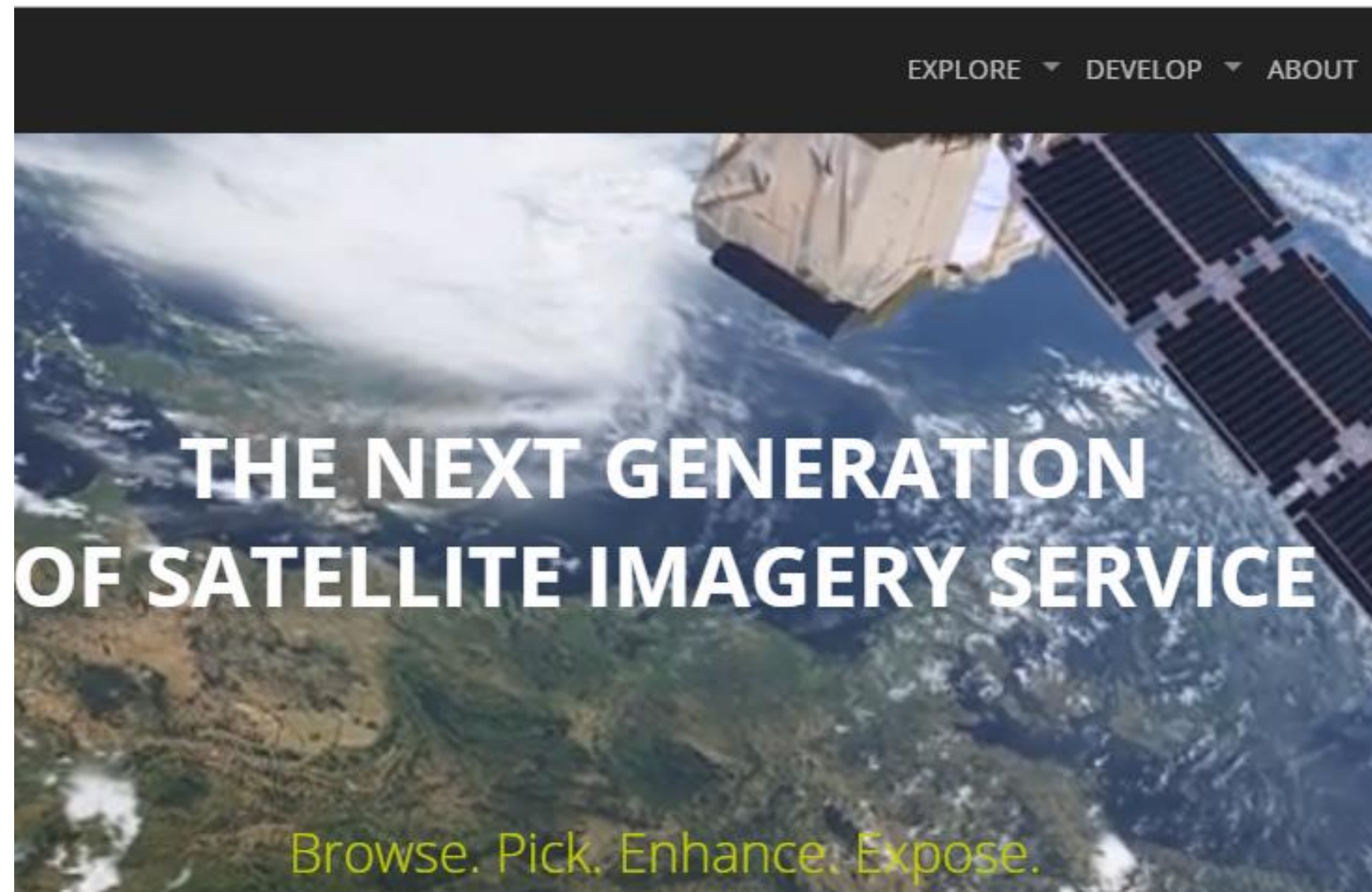


Dados de Entrada

Como obter imagens do Sentinel-2



THE
DEVELOPER'S
CONFERENCE



```
EOPatch(  
  data: {  
    BANDS: _FileLoader('./data/eopatch_new/eopatch_@_row-@_col-@/data/BANDS.npy.gz')  
    CLP: _FileLoader('./data/eopatch_new/eopatch_@_row-@_col-@/data/CLP.npy.gz')  
    NDVI: _FileLoader('./data/eopatch_new/eopatch_@_row-@_col-@/data/NDVI.npy.gz')  
    NDWI: _FileLoader('./data/eopatch_new/eopatch_@_row-@_col-@/data/NDWI.npy.gz')  
    NORM: _FileLoader('./data/eopatch_new/eopatch_@_row-@_col-@/data/NORM.npy.gz')  
  }  
  mask: {  
    CLM: _FileLoader('./data/eopatch_new/eopatch_@_row-@_col-@/mask/CLM.npy.gz')  
    IS_DATA: _FileLoader('./data/eopatch_new/eopatch_@_row-@_col-@/mask/IS_DATA.npy.gz')  
    IS_VALID: _FileLoader('./data/eopatch_new/eopatch_@_row-@_col-@/mask/IS_VALID.npy.gz')  
  }  
  scalar: {}  
  label: {}  
  vector: {}  
  data_timeless: {}  
  mask_timeless: {  
    LULC: _FileLoader('./data/eopatch_new/eopatch_@_row-@_col-@/mask_timeless/LULC.npy.gz')  
    VALID_COUNT: _FileLoader('./data/eopatch_new/eopatch_@_row-@_col-@/mask_timeless/VALID_C  
  }  
  scalar_timeless: {}  
  label_timeless: {}  
  vector_timeless: {}  
  meta_info: {  
    maxcc: 0.8  
    service_type: 'wcs'  
    size_x: '10m'  
    size_y: '10m'  
    time_difference: datetime.timedelta(-1, 86399)  
    time_interval: ['2017-07-01', '2017-12-30']  
  }  
  bbox: BBox((-53.29154208900681, -29.463377671276096), (-53.28483083909287, -29.453525015  
  timestamp: [datetime.datetime(2017, 7, 2, 13, 38, 21), ..., datetime.datetime(2017, 12, 2  
)
```

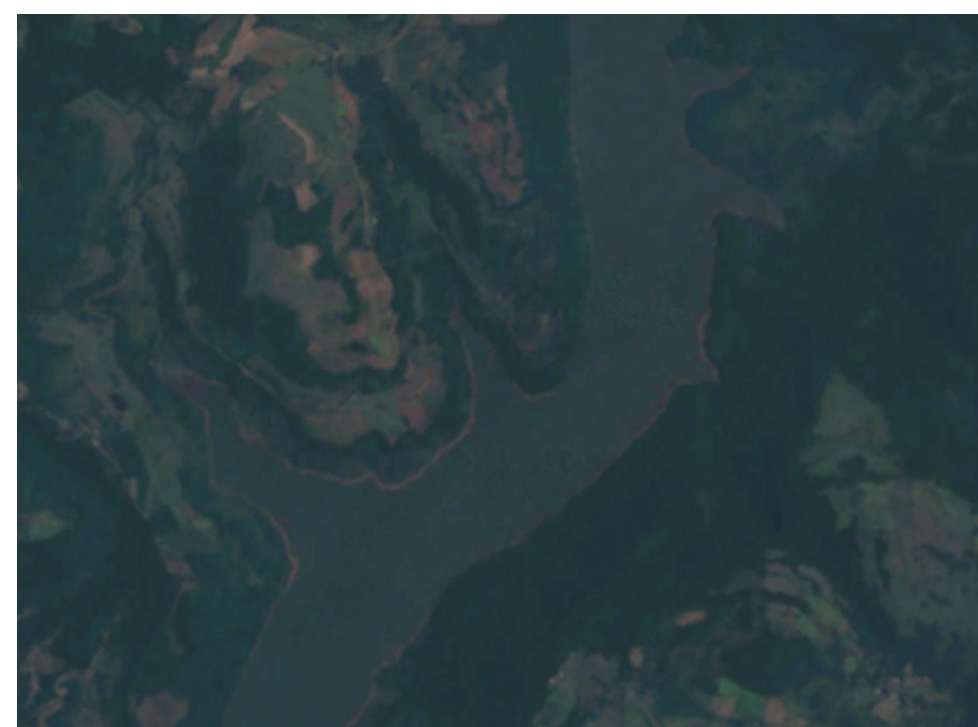
Informações Sentinel-2



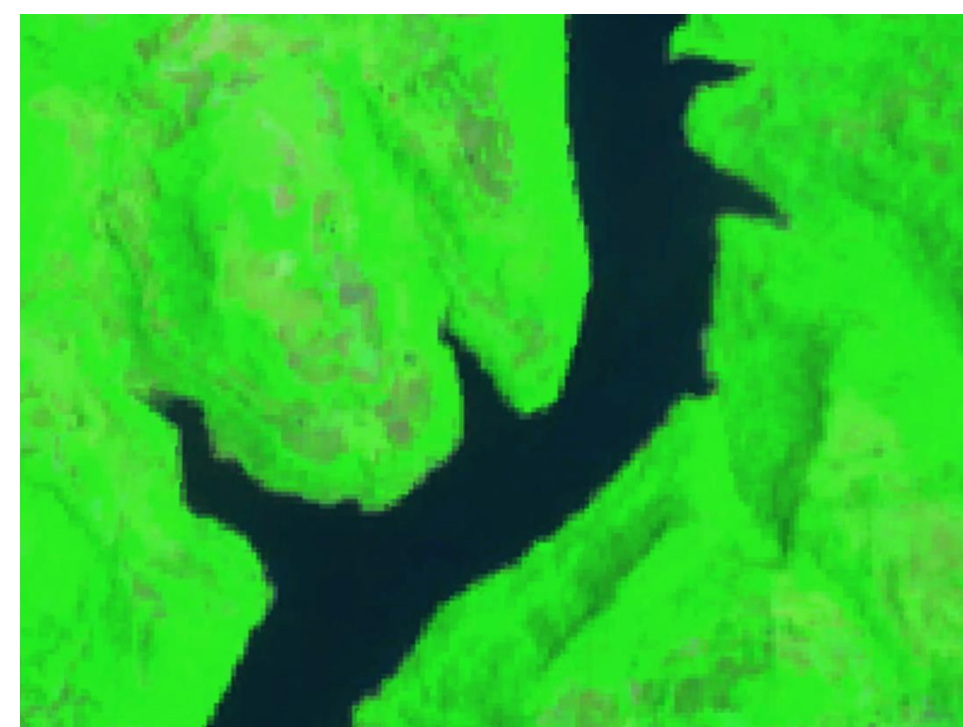
THE
DEVELOPER'S
CONFERENCE

- Imagens do satélite Sentinel-2 obtidas através do Sentinel Hub e EOlearn

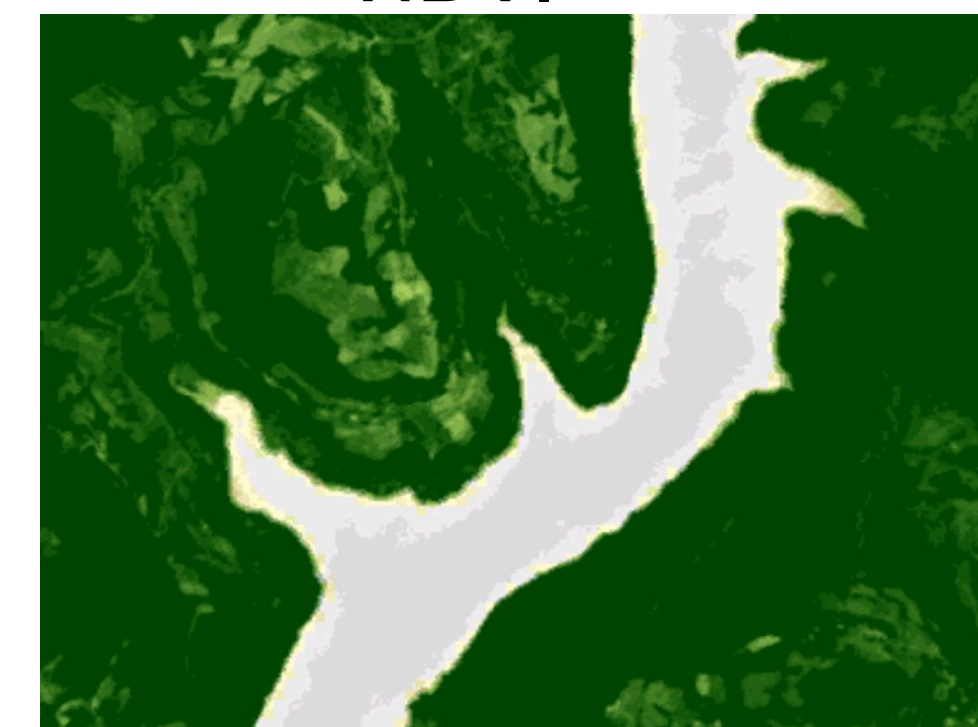
TRUE COLOR



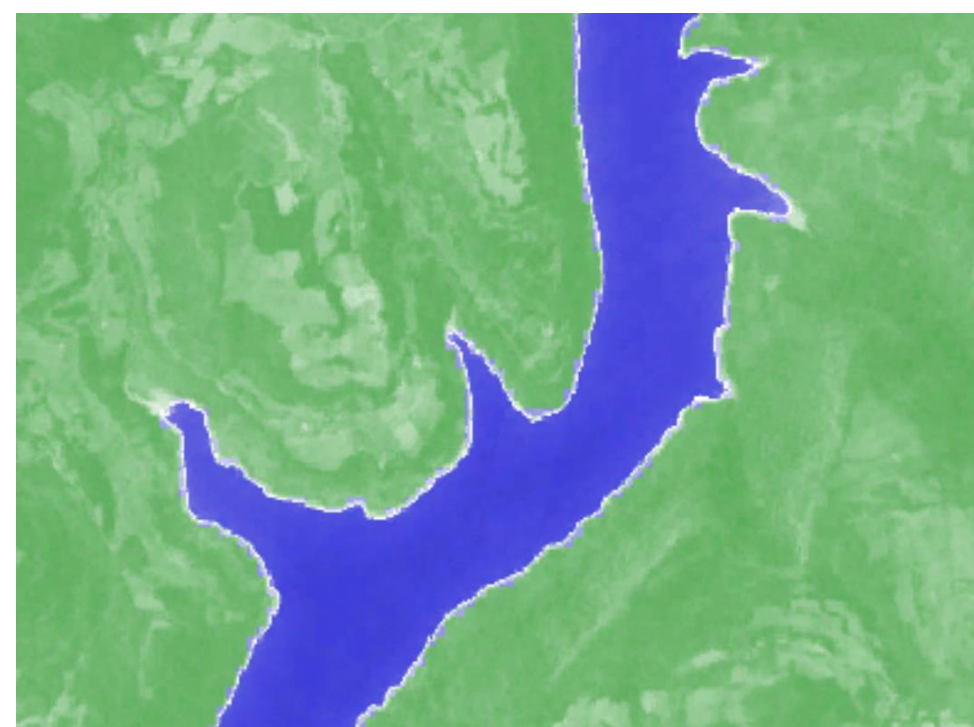
SWIR



NDVI



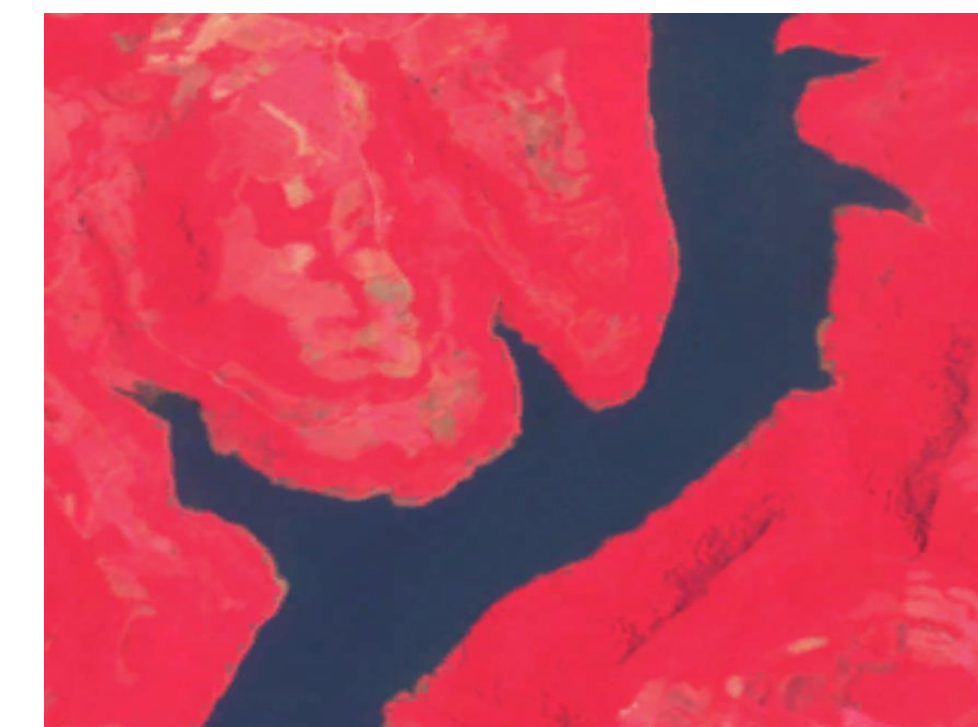
NDWI



NDSI



FALSE COLOR


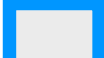





Anotações



THE
DEVELOPER'S
CONFERENCE

Definimos polígonos para determinar áreas conhecidas gerando uma base de dados para treinamento do algoritmo.

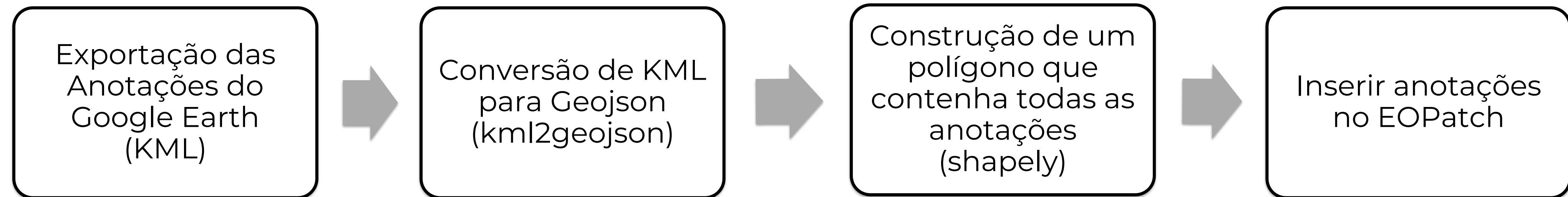
-  Vegetação campestre
-  Floresta em estágio inicial
-  Floresta em estágio médio e avançado
-  Sombra
-  Rio



Anotações



THE
DEVELOPER'S
CONFERENCE

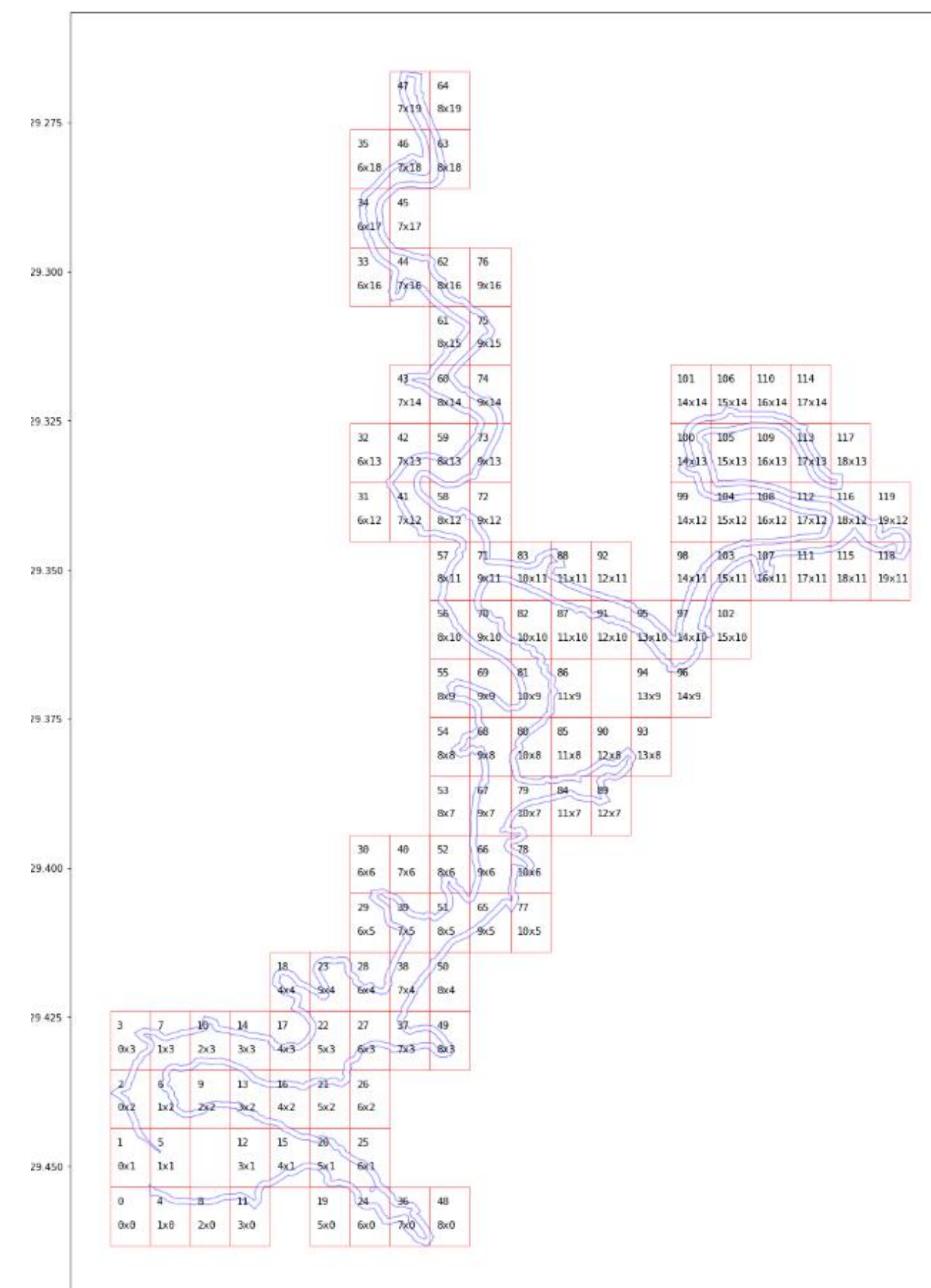


Pré-processamento dos Dados

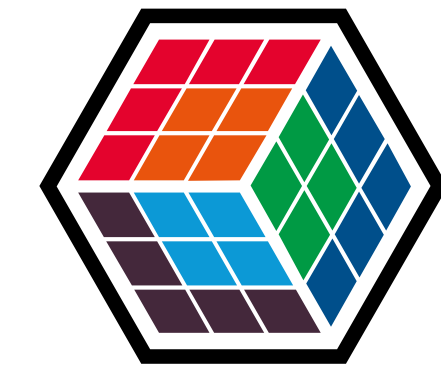


THE
DEVELOPER'S
CONFERENCE

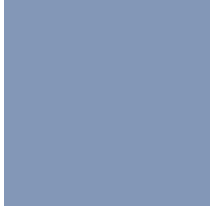


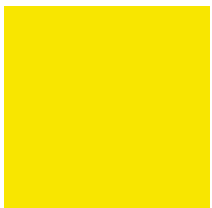

1. Secção da Área de interesse para aquisição das imagens de satélite
2. Filtragem de Nuvens
3. Composição de Imagem com Média dos Pixels.
4. Amostragem dos Pixels
5. Separação das seções para treinamento e teste
6. Treinamento e teste do modelo (LightGBM)
7. Exportação dos resultados em GeoTIFF (GDAL)

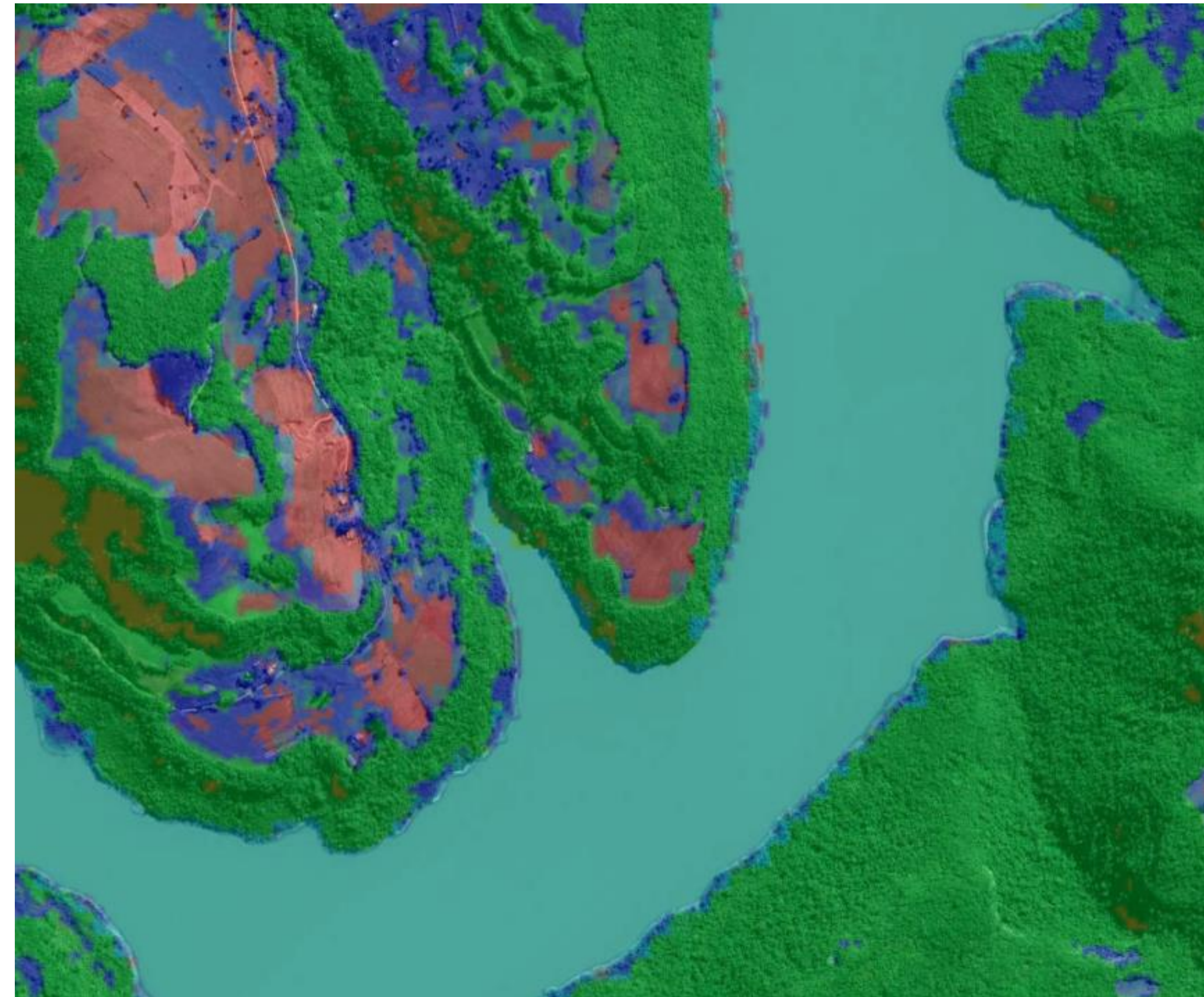


Resultado



THE
DEVELOPER'S
CONFERENCE

-  Vegetação campestre
-  Floresta em estágio inicial
-  Floresta em estágio médio e avançado
-  Sombra
-  Rio



Avaliação do Modelo



THE
DEVELOPER'S
CONFERENCE

Avaliação do Modelo

Desempenho final do modelo:

- Acurácia na classificação: **98.2%**
- F1-score para classificação: **98.2%**

Classe	F1	Recall	Precisão
Vegetação Campestre	65,9	66,3	65,6
Vegetação Inicial	73,7	73,1	74,3
Vegetação Média/Avançada	98,4	99,2	97,5
Sombra	93,5	87,8	100,0
Rio	99,9	100,0	99,9

Precisão: Número de vezes que a classe foi predita corretamente dado que aquela classe foi predita.
Recall: Número de vezes que uma classe foi predita corretamente dividido pelo número de vezes que a classe aparece nos dados de teste.
F1: Média harmônica entre precisão e recall, indicando a performance do classificador com apenas um indicador.

MATRIZ DE CONFUSÃO

Vegetação Campestre	0.66	0.32	0.02	0.00	0.00
Vegetação Inicial	0.22	0.73	0.04	0.00	0.01
Vegetação Med. / Avanc.	0.00	0.01	0.99	0.00	0.00
Sombra	0.00	0.00	0.11	0.88	0.01
Rio	0.00	0.00	0.00	0.00	1.00
	Vegetação Campestre	Vegetação Inicial	Vegetação Med. / Avanc.	Sombra	Rio

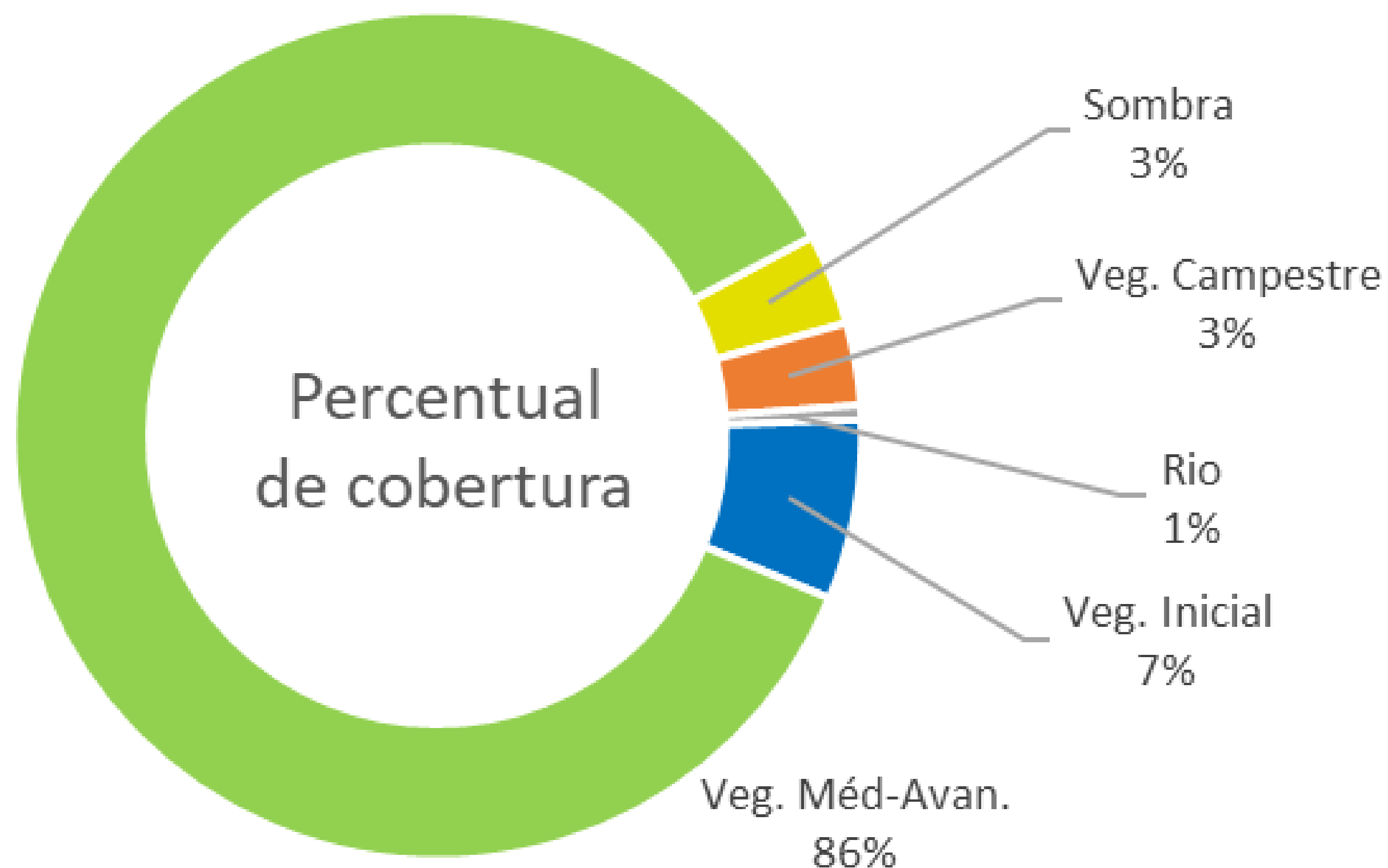
PREDICTED (LGBM)

Relatório Final



THE
DEVELOPER'S
CONFERENCE

A partir do total da área de interesse e das classes categorizadas, calcula:



Área em m² de cada classe:

Classes	Área m ²
Veg. <u>Inicial</u>	867.844,37
Veg. <u>Med-Avancada</u>	8.921.517,46
Veg. <u>Campestre</u>	262.097,24
Sombra	127.853,23
Rio	454.452,75



THE
DEVELOPER'S
CONFERENCE

Obrigado!



THE DEVELOPER'S CONFERENCE