



Potencial de técnicas de mineração de dados para o mapeamento de áreas cafeeiras

Cesare Di Girolamo Neto – 130338

São José dos Campos
Junho, 2014



Introdução:

- Brasil é o maior produtor de café do mundo (37%) → US\$ 6 bilhões/ano.

- Metodologias tradicionais de mapeamento do café:
 - Alto custo
 - Dificuldade na obtenção de dados
 - Lentas quando comparadas a métodos computacionais

- Técnicas de mineração de dados:
 - Alta capacidade de reconhecimento de padrões
 - Baixo custo
 - Pode ser aplicada a imagens de satélite



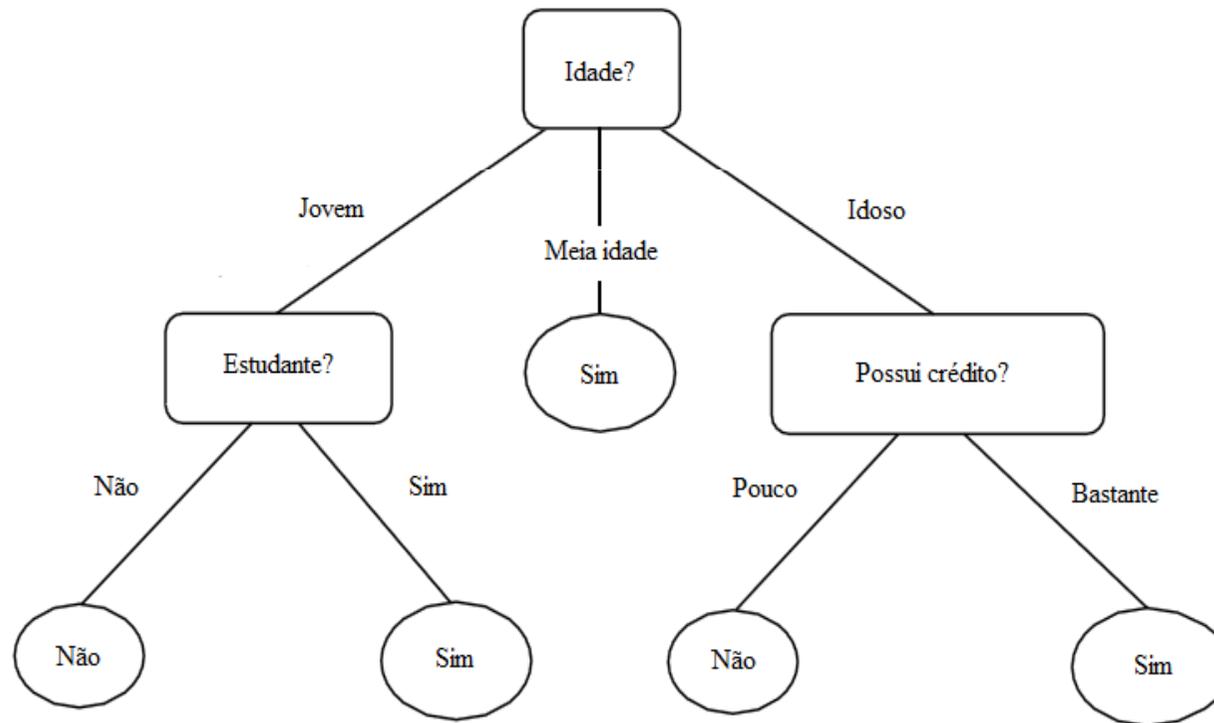
Objetivo:

Este trabalho teve como objetivo avaliar o potencial de técnicas de mineração de dados para classificação automática de áreas cafeeiras utilizando imagens do LANDSAT-8 para o município de Machado/MG.



Técnicas de mineração de dados:

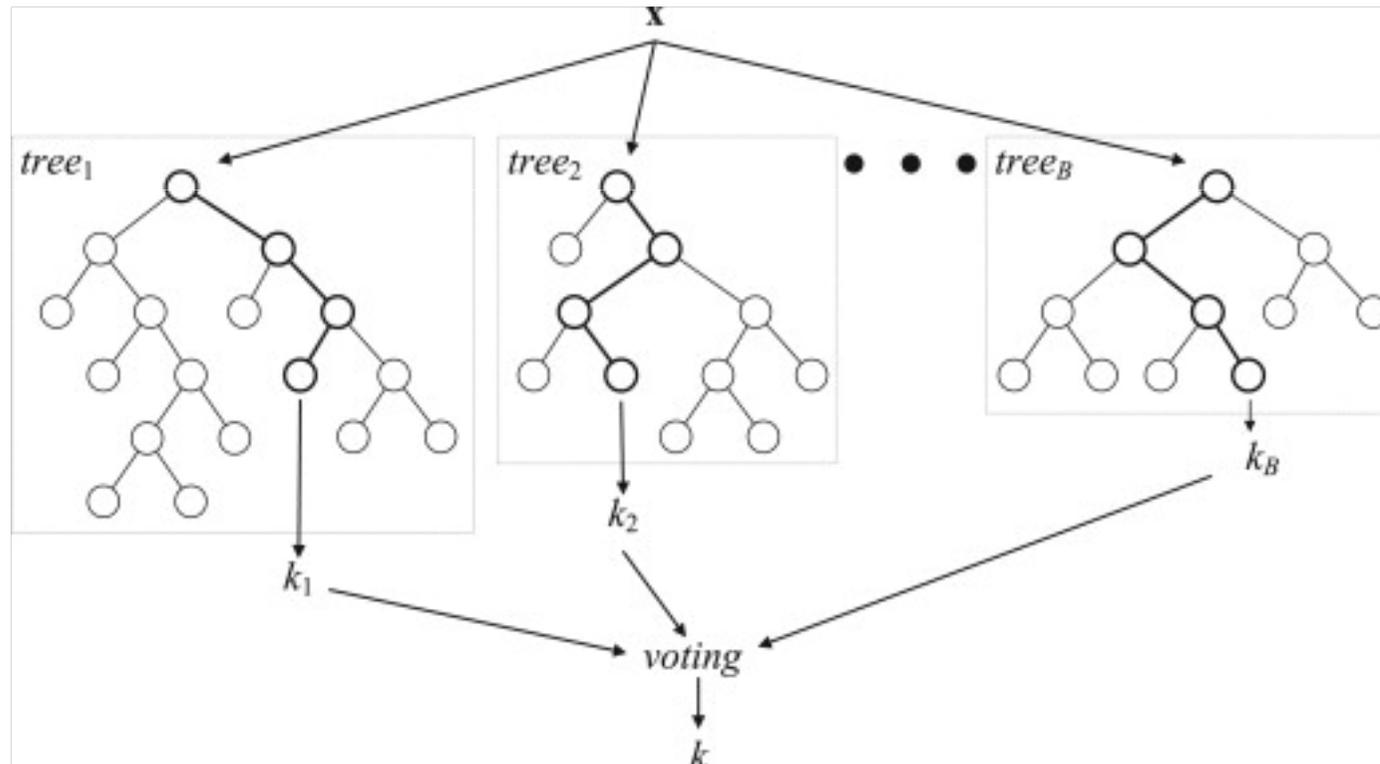
- Árvores de decisão





Técnicas de mineração de dados:

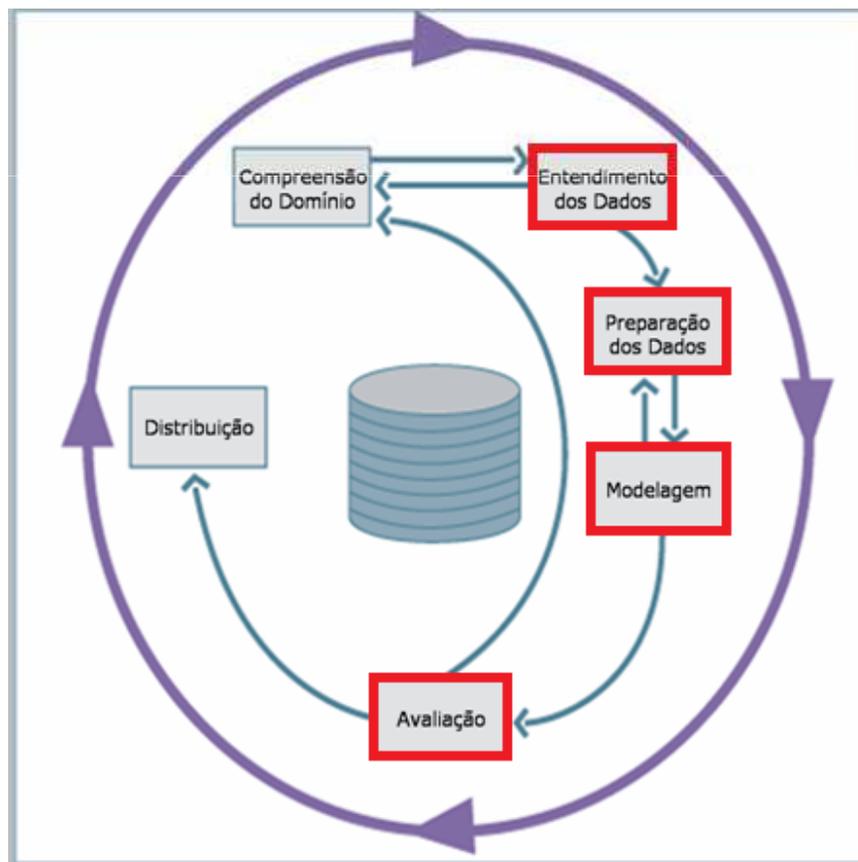
- Random Forest

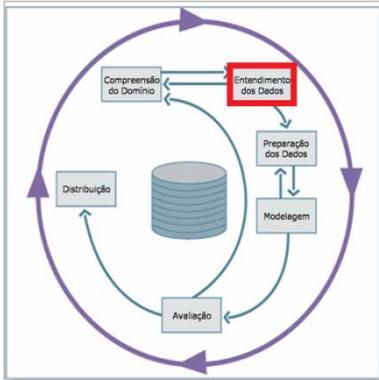




Metodologia:

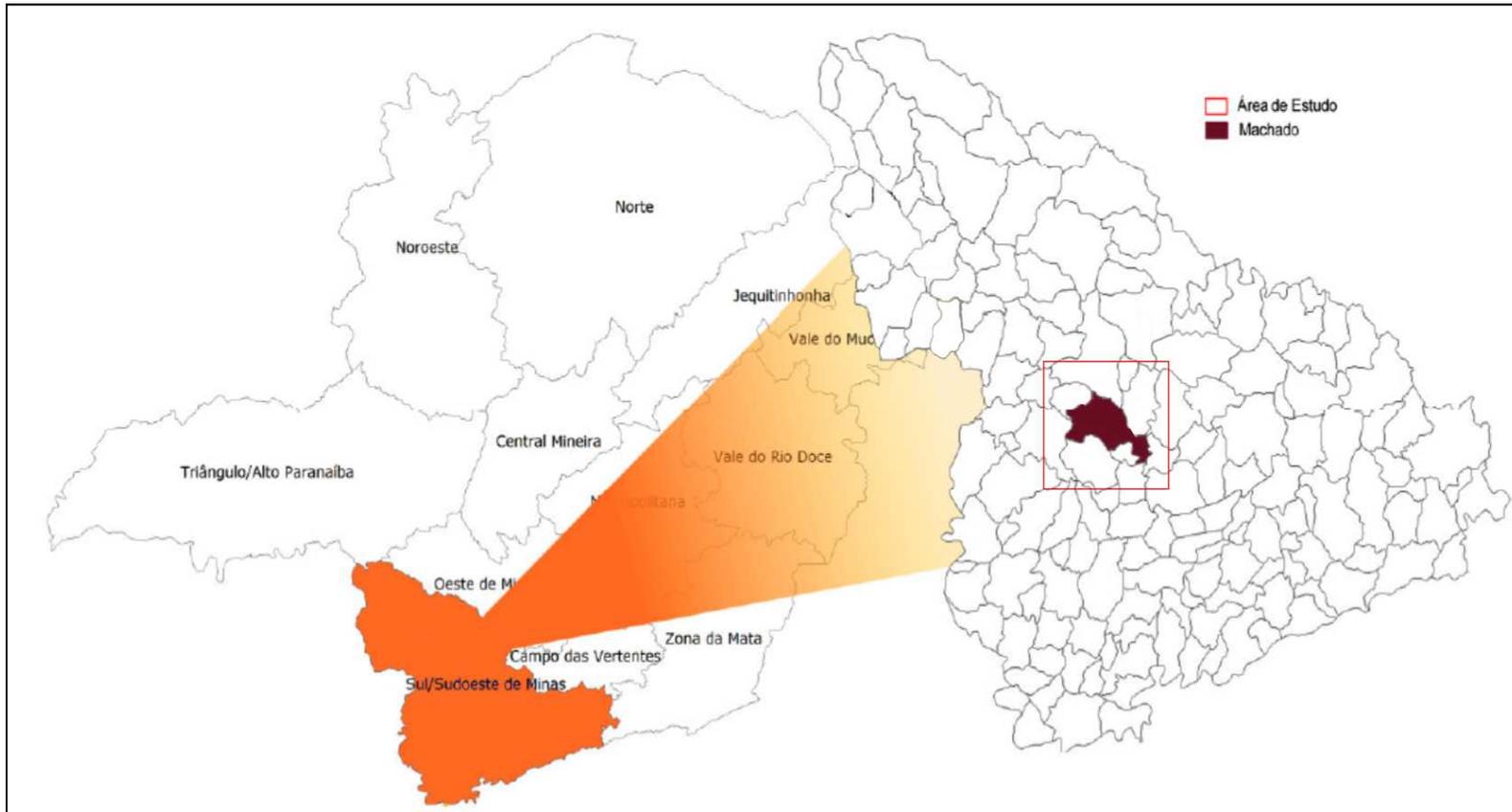
- CRISP-DM (CHAPMAN et al., 2000)

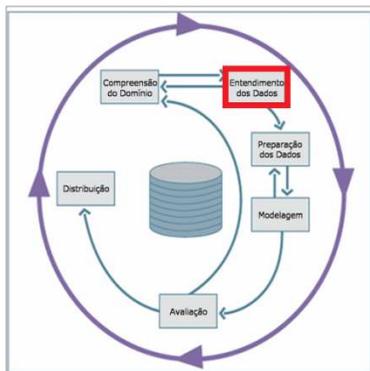




Fase de entendimento dos dados:

- Área de estudo: Município de Machado/MG.





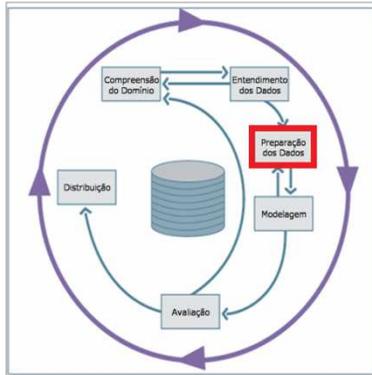
Fase de entendimento dos dados:

- Imagem LANDSAT-8 com 11 bandas datada de 08/03/2014.
- Imagem temática com classificação manual realizada pelo Prof. Moreira (CAFESAT).
- Shapefiles com máscaras de drenagem e de municípios de Minas Gerais (Banco de dados DPI/INPE).

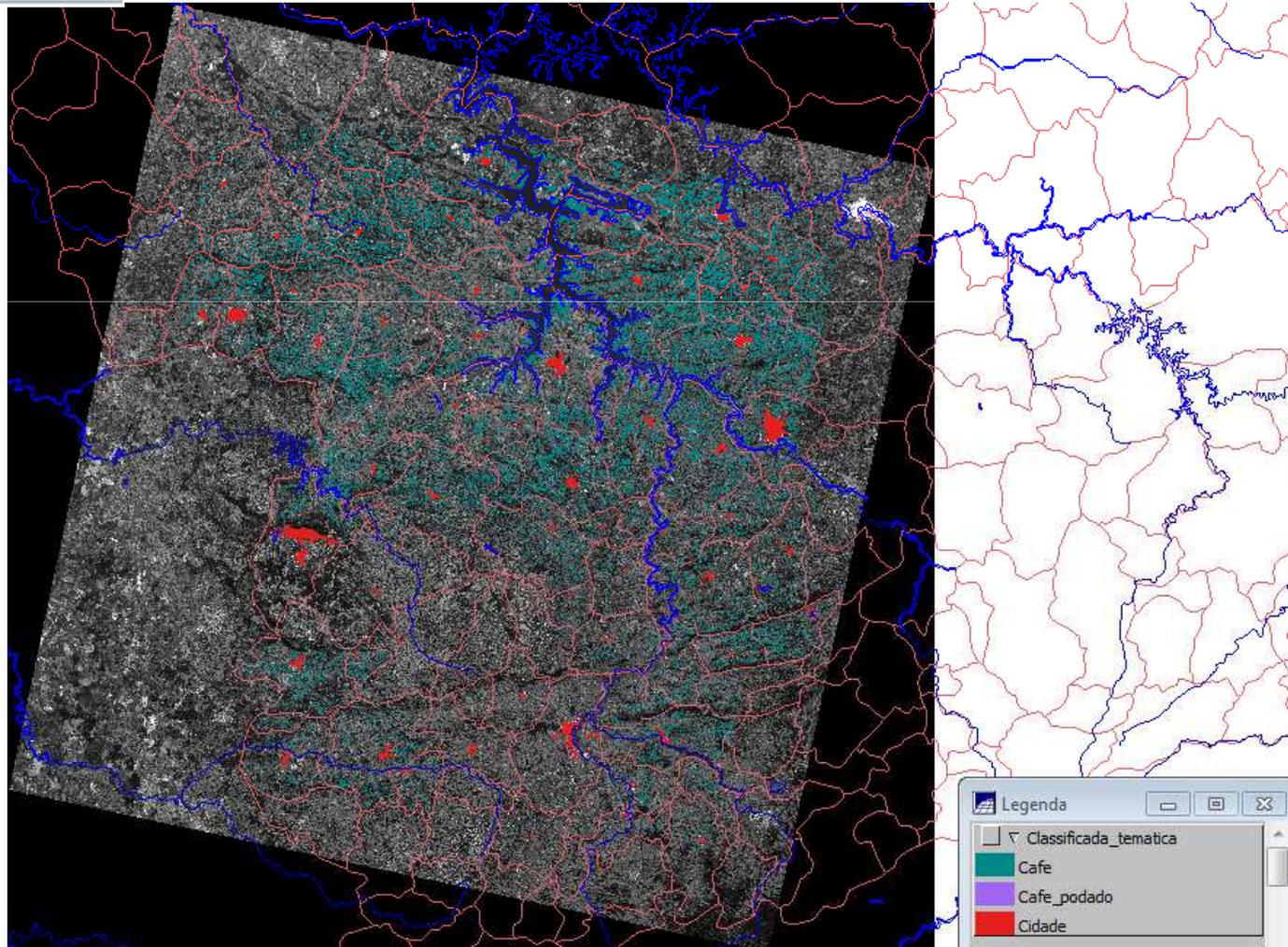


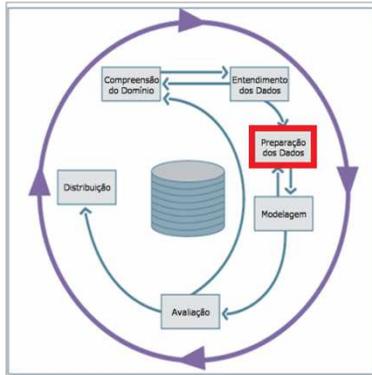
SPRING

Sistema de Processamento de
Informações Georeferenciadas

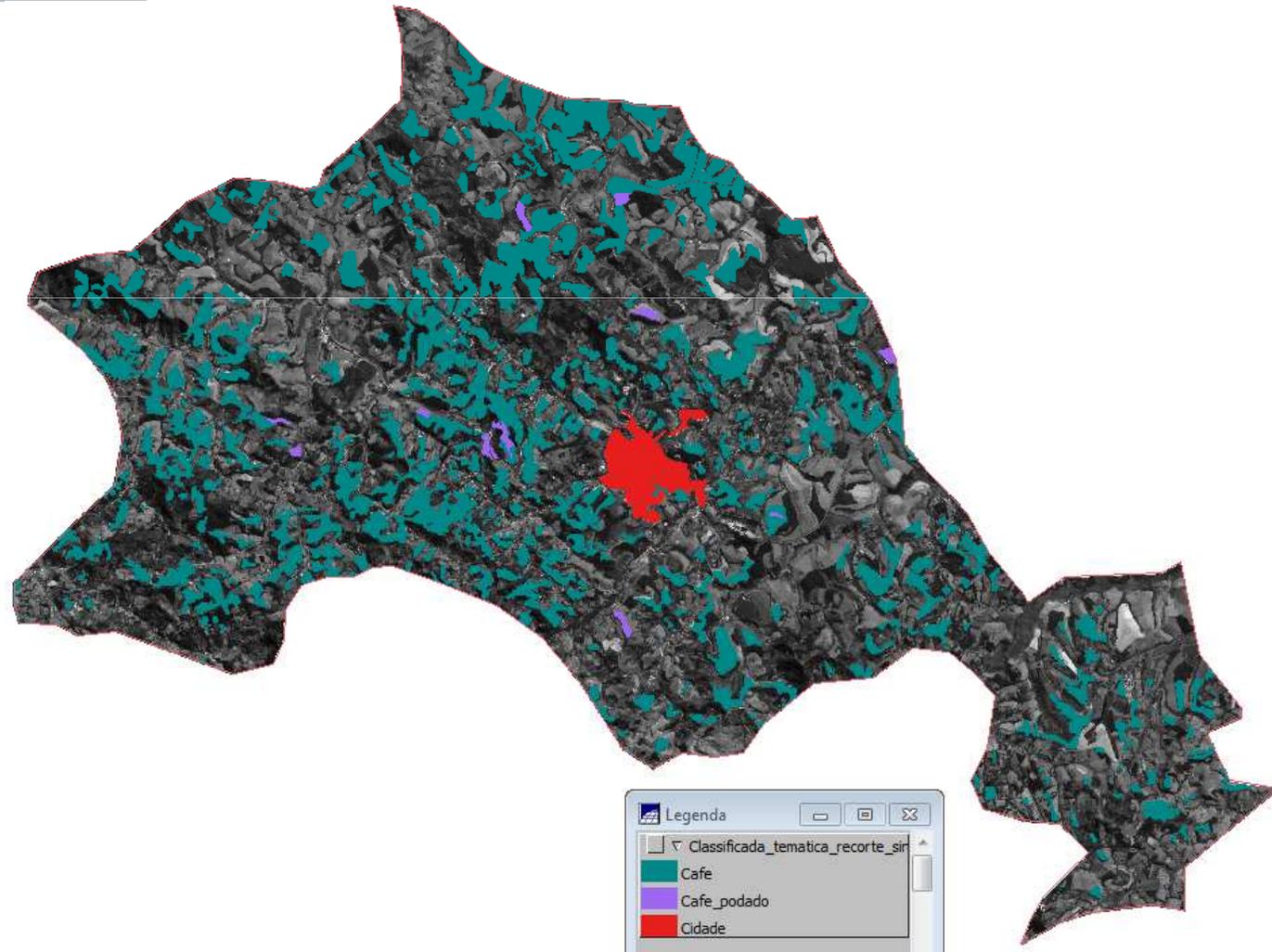


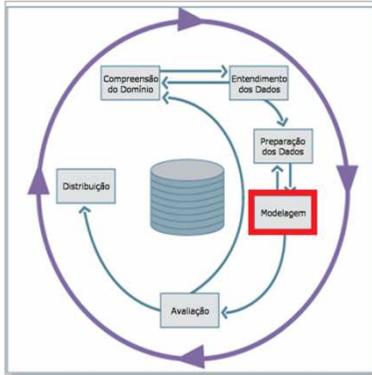
Fase de preparação dos dados:





Fase de preparação dos dados:

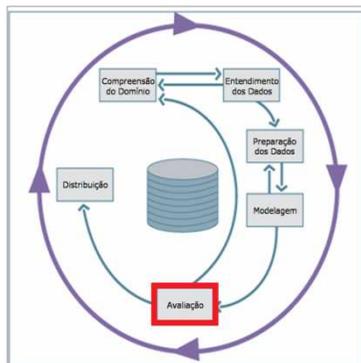




Fase de modelagem:

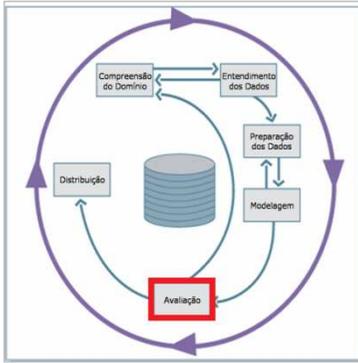
- Níveis de cinza extraídos (.csv) por meio de Scripts em C++.
- Conjunto preparado para a modelagem.
 - 11 Atributos
 - 645.785 registros
- Amostragem de 25% para treinamento do modelo.





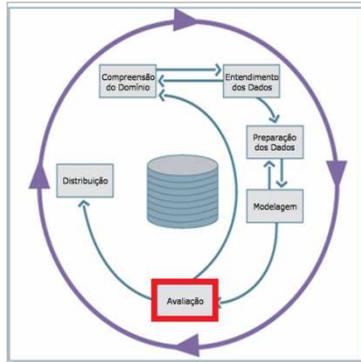
Resultados:

	Árvores de decisão (variando o número de registros por folha).					
	30000	15000	7500	1500	750	250
Taxa de acerto	68,73%	75,96%	76,10%	75,93%	77,47%	76,90%
Índice Kappa	0,30	0,36	0,38	0,40	0,42	0,42
N. de folhas	3	4	7	27	39	118



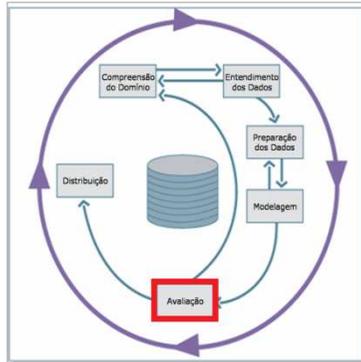
Resultados:

	Random Forest (variando o número de árvores na floresta)			
	5	10	25	100
Taxa de acerto	81,70%	81,00%	83,97%	84,13%
Índice Kappa	0,54	0,54	0,60	0,60

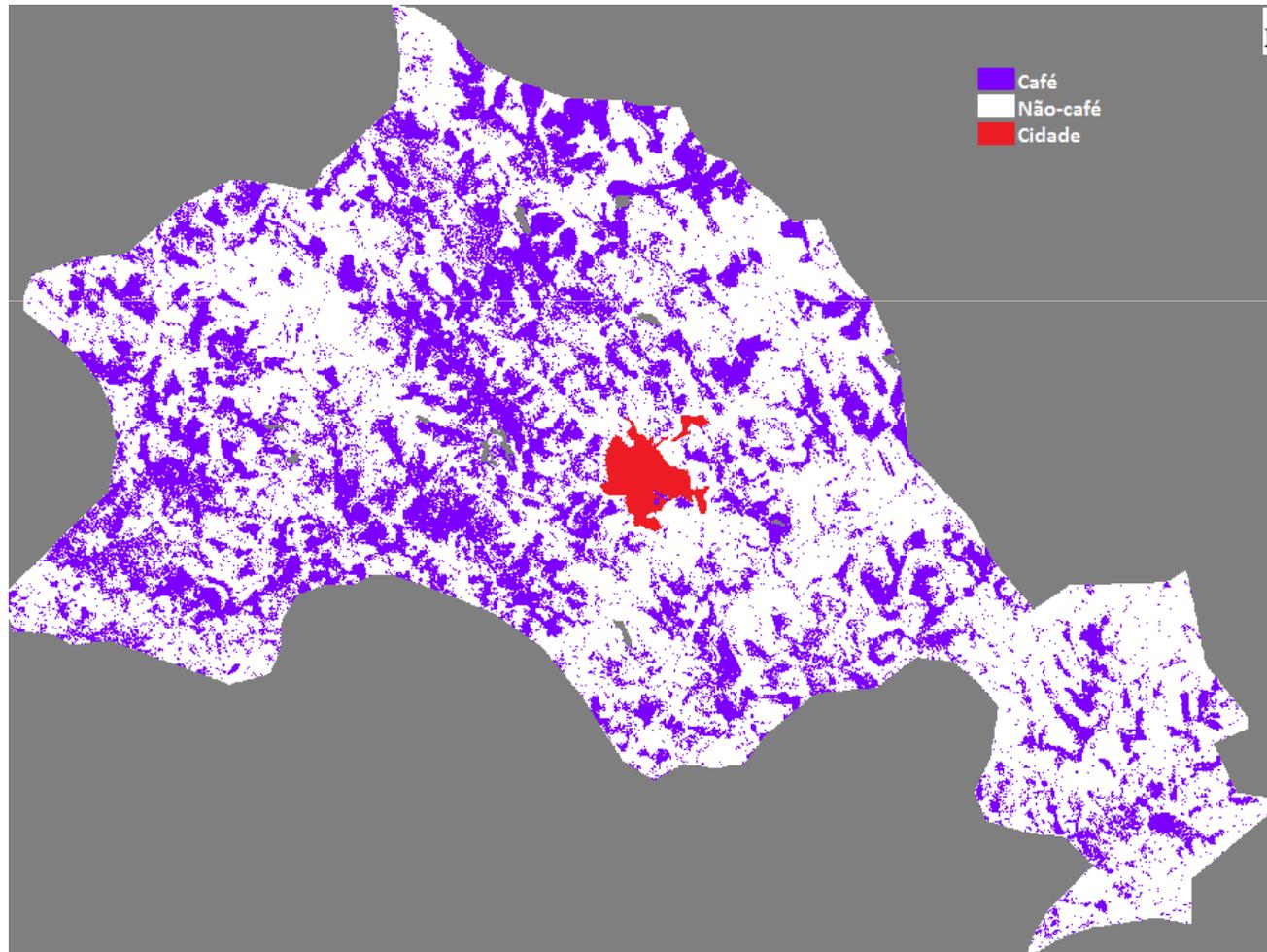


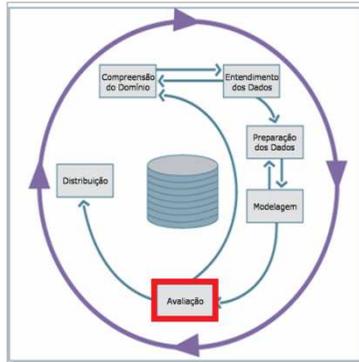
Resultados:

- Random Forest melhores que árvores de decisão e compatíveis com classificadores desenvolvidos por outros autores.
- Métodos de seleção de atributos foram aplicados mas não melhoraram o desempenho dos modelos, apenas reduziram seu custo computacional (25%)
 - Bandas 7, 9 e 10 descartadas.

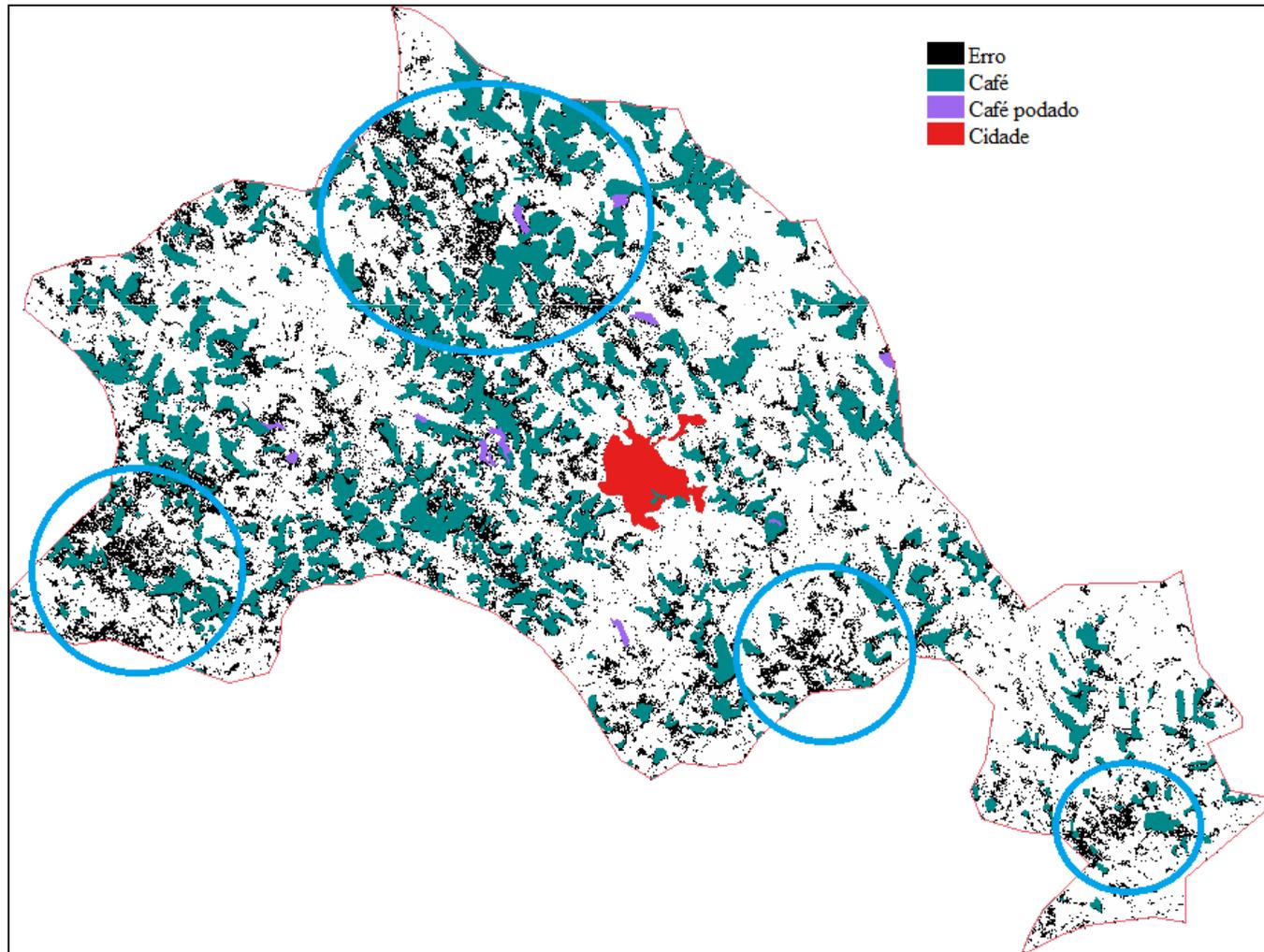


Resultados:





Resultados:





Conclusões:

- Uma classificação de imagem baseada em Random Forest contendo apenas características espectrais deve ser utilizada como suporte à outras classificações.
- A diferenciação de áreas com relevo íngreme e de outras culturas agrícolas pode melhorar no desempenho dos classificadores.
- Inclusão de novos atributos no conjunto de dados como altitude, orientação de vertente e declive poderia melhorar a classificação.