

Projeto - Reconhecimento de Padrões

1. (10 points) Considere um problema de *Reconhecimento de Padrões* cujo objetivo é reconhecer o padrão de produtividade de uma determinada área cultivada. Para tal, tem-se o resultado da Análise do Solo (`baseSolo.csv`) e Mapa de colheita (`baseColheita.csv`) de determinada área cultivada. O atributo que designa a produtividade é o “Rendimento seco”.

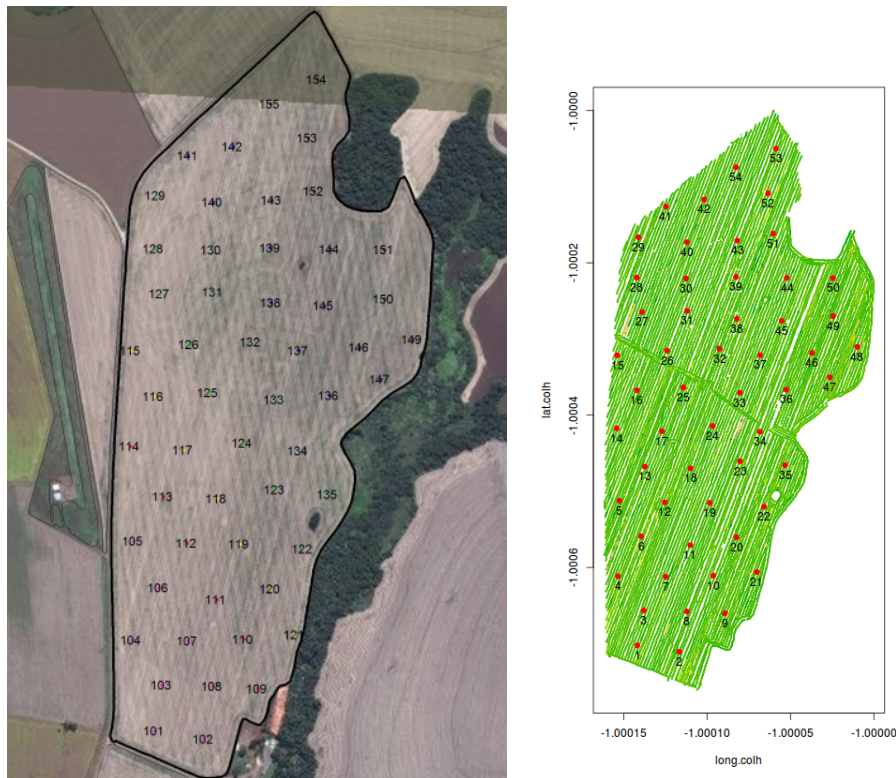


Table 1: [Esquerda] - Mapa do terreno com pontos de análise de solo e [Direita] plot com pontos de colheita de solo (sub-amostrados) e pontos de análise de solo.

- **Base de Análise de Solo:**

- Atributos:
- Instâncias: 54 pontos de solo analisados.

- **Base de Resultado de Colheita:**

- Atributos: "Latitude", "Longitude", "Elevacao", "Distancia", "Duracao", "Largura", "Titulo", "Registro.de.horario", "Qualidade.GPS", "Umidade", "Fluxo.Corn", "Section.ID", "Acoplado", "Velocidade", "Rendimento.de.massa", "Rendimento.umido", "Rendimento.seco"
- Instâncias: 110.196 pontos de colheita.

É esperado que seja desenvolvido um relatório com os processos desenvolvidos para a detecção dos padrões de alta produtividade. Considere a criação de fluxogramas para explicitar as ações utilizadas e se possível, plote os resultados por meio de mapas. Utilize as seguintes questões como guia para a condução dos experimentos:

1. Quais as áreas mais produtivas?
2. Qual a composição mais frequente da área mais produtiva?
3. Qual o nutriente que mais corresponde positivamente nas áreas mais produtivas?
4. Qual a área menos produtiva?
5. Qual a composição mais frequente da área menos produtiva?

Sugestões:

- Por ser um base ser uma base real, está repleta de outliers e ruído.
- As bases são desbalanceadas, considere o uso de subamostragem na base majoritária.
- Crie um dataset com apenas as informações necessárias através da união por meio de atributos compartilhados.
- Relaxe o problema. Encare ele como um problema binário e depois multi-classes com poucos rótulos.
- Utilize métricas de avaliação para analisar a qualidade dos dados e processos desenvolvidos.