

DESAFIO TREEVIA

ANA CAROLINA

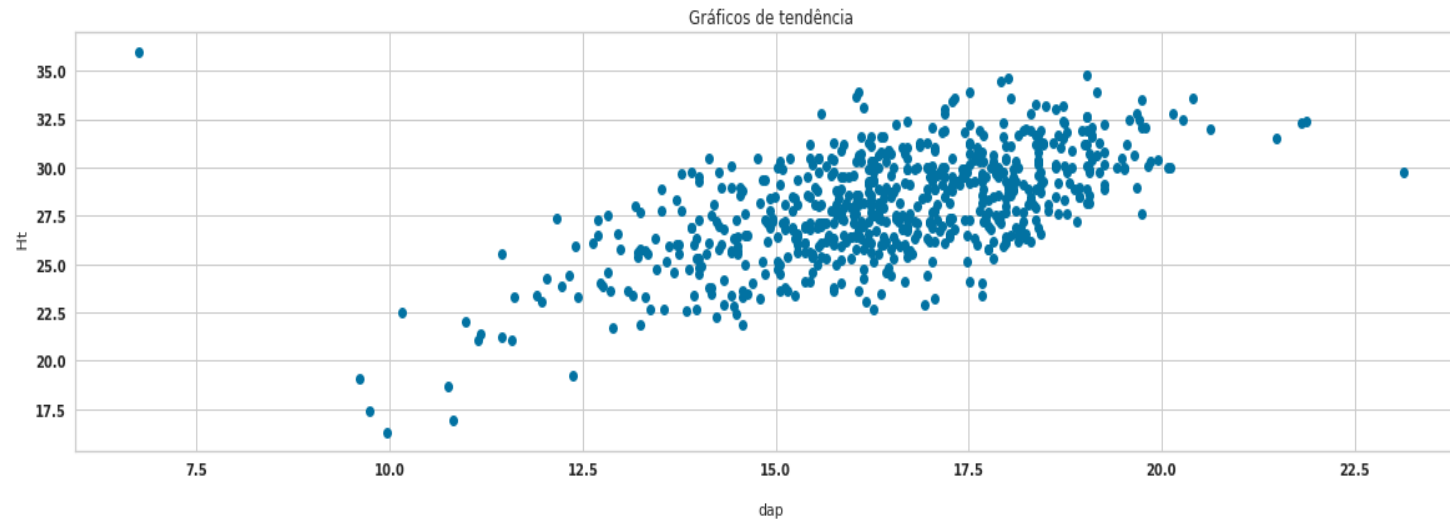
<https://github.com/ana181084>

Apresentação do desafio

- A base de dados em anexo foi obtida por meio da realização de um Inventário Florestal. Pré Corte (IPC), em fazenda localizada no Estado de Mato Grosso do Sul (MS). Os dados obtidos em campo foram auditados e aprovados pelo contratante.
- Você foi escolhido(a) para realizar o processamento deste inventário juntamente com a sua defesa técnica. As variáveis de interesse são: Volume Total com Casca (VTCC) e Volume Comercial com Casca até 5cm da ponta fina (VCOM5cm). Apresente erro e intervalo de confiança com nível de significância a 95%.

Modelagem da altura:

- Plantio com 8 anos de idade ;
- Calculado o DAP e $\text{Invdap}(1/\text{dap})$;
- Calculado o Ln da altura total;
- As Falhas foram retiradas para ajuste do modelo;



- Retirado o outlier ($\text{dap} < 7.5$ e > 21)

Modelagem da altura:

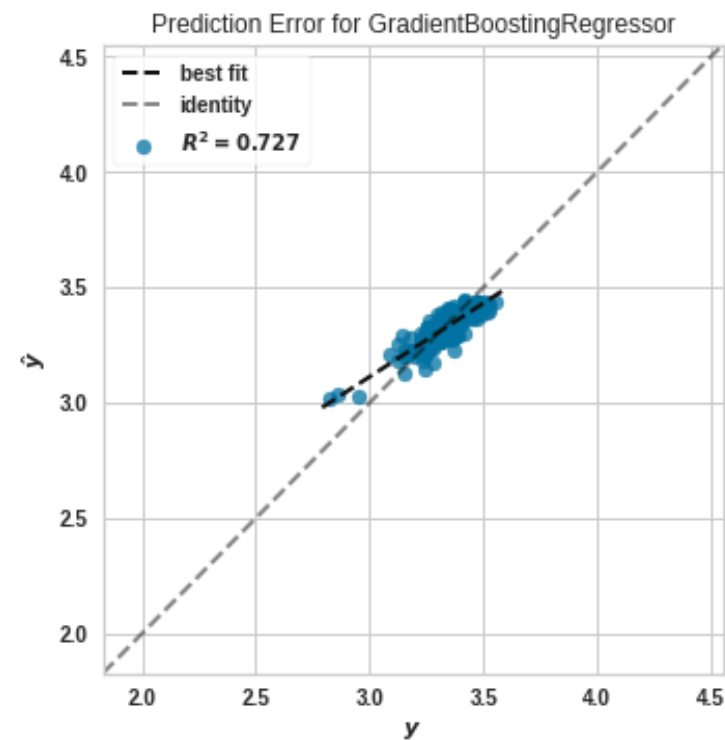
PYCARET (Escolha do modelo menor 'RMSE')

| | Data Type |
|-------------|-------------|
| CoordenadaX | Numeric |
| CoordenadaY | Numeric |
| Categoria1 | Categorical |
| Fuste | Categorical |
| 1/dap | Numeric |
| InHt | Label |

| | Model | MAE | MSE | RMSE | R2 | RMSLE | MAPE | TT (Sec) |
|----------|---------------------------------|--------|--------|--------|----------|--------|--------|----------|
| gbr | Gradient Boosting Regressor | 0.0448 | 0.0035 | 0.0588 | 0.6515 | 0.0137 | 0.0138 | 0.059 |
| rf | Random Forest Regressor | 0.0467 | 0.0038 | 0.0608 | 0.6300 | 0.0142 | 0.0141 | 0.456 |
| lightgbm | Light Gradient Boosting Machine | 0.0472 | 0.0039 | 0.0618 | 0.6185 | 0.0145 | 0.0143 | 0.062 |
| et | Extra Trees Regressor | 0.0523 | 0.0047 | 0.0679 | 0.5345 | 0.0159 | 0.0158 | 0.526 |
| ada | AdaBoost Regressor | 0.0549 | 0.0049 | 0.0691 | 0.5140 | 0.0161 | 0.0166 | 0.054 |
| knn | K Neighbors Regressor | 0.0529 | 0.0056 | 0.0730 | 0.4809 | 0.0171 | 0.0161 | 0.060 |
| lar | Least Angle Regression | 0.0620 | 0.0061 | 0.0778 | 0.3843 | 0.0181 | 0.0187 | 0.013 |
| br | Bayesian Ridge | 0.0620 | 0.0061 | 0.0779 | 0.3843 | 0.0181 | 0.0188 | 0.013 |
| omp | Orthogonal Matching Pursuit | 0.0628 | 0.0062 | 0.0783 | 0.3766 | 0.0182 | 0.0190 | 0.012 |
| dt | Decision Tree Regressor | 0.0622 | 0.0064 | 0.0794 | 0.3562 | 0.0186 | 0.0188 | 0.017 |
| ridge | Ridge Regression | 0.0795 | 0.0101 | 0.0993 | 0.0310 | 0.0232 | 0.0241 | 0.012 |
| lr | Linear Regression | 0.0825 | 0.0108 | 0.1027 | -0.0321 | 0.0240 | 0.0250 | 0.296 |
| lasso | Lasso Regression | 0.0825 | 0.0108 | 0.1027 | -0.0321 | 0.0240 | 0.0250 | 0.016 |
| en | Elastic Net | 0.0825 | 0.0108 | 0.1027 | -0.0321 | 0.0240 | 0.0250 | 0.014 |
| llar | Lasso Least Angle Regression | 0.0830 | 0.0109 | 0.1030 | -0.0392 | 0.0240 | 0.0252 | 0.011 |
| dummy | Dummy Regressor | 0.0830 | 0.0109 | 0.1030 | -0.0392 | 0.0240 | 0.0252 | 0.010 |
| huber | Huber Regressor | 0.2638 | 0.2795 | 0.5215 | -28.1113 | 0.1529 | 0.0803 | 0.019 |
| par | Passive Aggressive Regressor | 0.4504 | 0.5912 | 0.7421 | -62.3170 | 0.2232 | 0.1358 | 0.013 |

Modelagem da altura:

| | Model | MAE | MSE | RMSE | R2 | RMSLE | MAPE | TT (Sec) |
|-----|-----------------------------|--------|--------|--------|--------|--------|--------|----------|
| gbr | Gradient Boosting Regressor | 0.0448 | 0.0035 | 0.0586 | 0.6515 | 0.0137 | 0.0136 | 0.059 |



Modelagem da altura:

- Predição (Label):

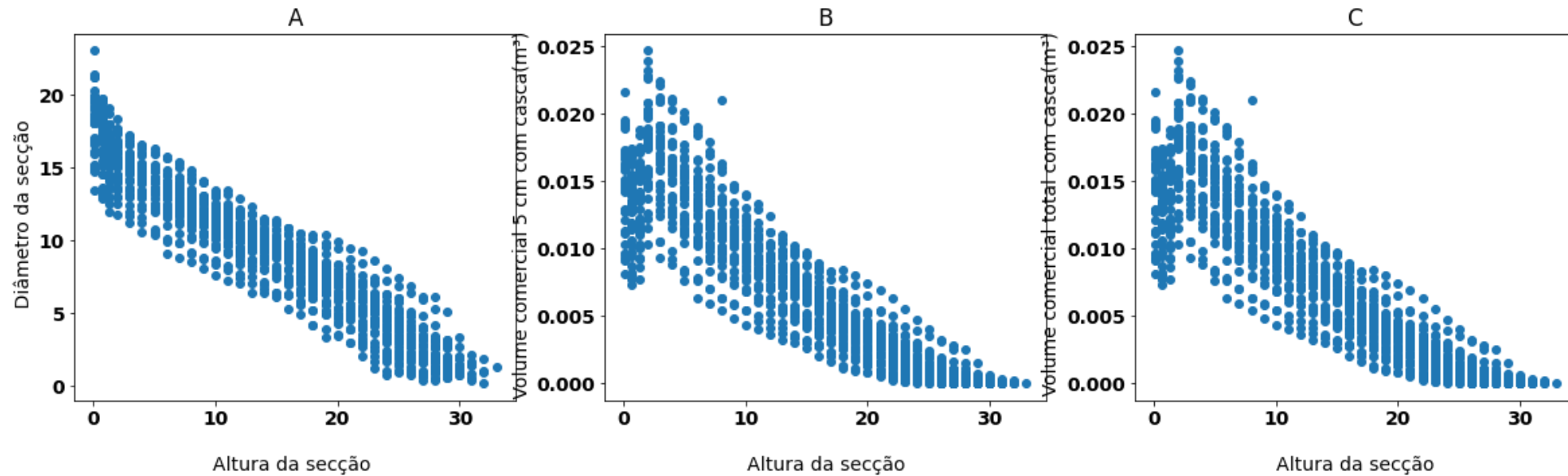
| | Talhao | Plantio | IdParcela | AreaParc | DataPlantio | DataColeta | Linha | Arvore | Fuste | Cap | Ht | Categoria1 | CoordenadaX | CoordenadaY | DAP | InHt | 1/dap | Label |
|------|--------|---------|------------|----------|-------------|------------|-------|--------|-------|------|------|------------|-------------|-------------|-----------|----------|----------|----------|
| 0 | T-8 | P-8 | IFC_12_001 | 400 | 01/05/2010 | 26/09/2018 | 1 | 1 | 1 | 57.5 | NaN | NORMAL | -5397725337 | -2049390493 | 18.302818 | NaN | 0.054636 | 3.405605 |
| 1 | T-8 | P-8 | IFC_12_008 | 400 | 01/05/2010 | 26/09/2018 | 2 | 5 | 1 | 58.0 | NaN | NORMAL | -5397815827 | -2049512442 | 18.461973 | NaN | 0.054165 | 3.413378 |
| 2 | T-8 | P-8 | IFC_12_008 | 400 | 01/05/2010 | 26/09/2018 | 2 | 6 | 1 | 40.4 | NaN | NORMAL | -5397815827 | -2049512442 | 12.859719 | NaN | 0.077762 | 3.258409 |
| 3 | T-8 | P-8 | IFC_12_008 | 400 | 01/05/2010 | 26/09/2018 | 3 | 1 | 1 | 56.5 | NaN | NORMAL | -5397815827 | -2049512442 | 17.984509 | NaN | 0.055603 | 3.398215 |
| 4 | T-8 | P-8 | IFC_12_008 | 400 | 01/05/2010 | 26/09/2018 | 3 | 2 | 1 | 49.2 | NaN | NORMAL | -5397815827 | -2049512442 | 15.660846 | NaN | 0.063854 | 3.366510 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 1691 | T-8 | P-8 | IFC_12_032 | 400 | 01/06/2010 | 26/09/2018 | 4 | 8 | 1 | 43.9 | 25.3 | NORMAL | -5397110174 | -2050786515 | 13.973804 | 3.230804 | 0.071562 | 3.238306 |
| 1692 | T-8 | P-8 | IFC_12_032 | 400 | 01/06/2010 | 26/09/2018 | 5 | 1 | 1 | 51.5 | 27.9 | NORMAL | -5397110174 | -2050786515 | 16.392959 | 3.328627 | 0.061002 | 3.312666 |
| 1693 | T-8 | P-8 | IFC_12_032 | 400 | 01/06/2010 | 26/09/2018 | 5 | 2 | 1 | 54.5 | 29.3 | NORMAL | -5397110174 | -2050786515 | 17.347889 | 3.377588 | 0.057644 | 3.326657 |
| 1694 | T-8 | P-8 | IFC_12_032 | 400 | 01/06/2010 | 26/09/2018 | 5 | 4 | 1 | 56.0 | 25.3 | NORMAL | -5397110174 | -2050786515 | 17.825354 | 3.230804 | 0.056100 | 3.326657 |
| 1695 | T-8 | P-8 | IFC_12_040 | 400 | 01/06/2010 | 26/09/2018 | 7 | 4 | 1 | 51.7 | NaN | NORMAL | -5397529703 | -2051046136 | 16.456621 | NaN | 0.060766 | 3.277652 |

- A altura não medida em campo foi estimado a partir do modelo ajustado

Modelagem do volume:

Cubagem de árvores com 2 anos de idade

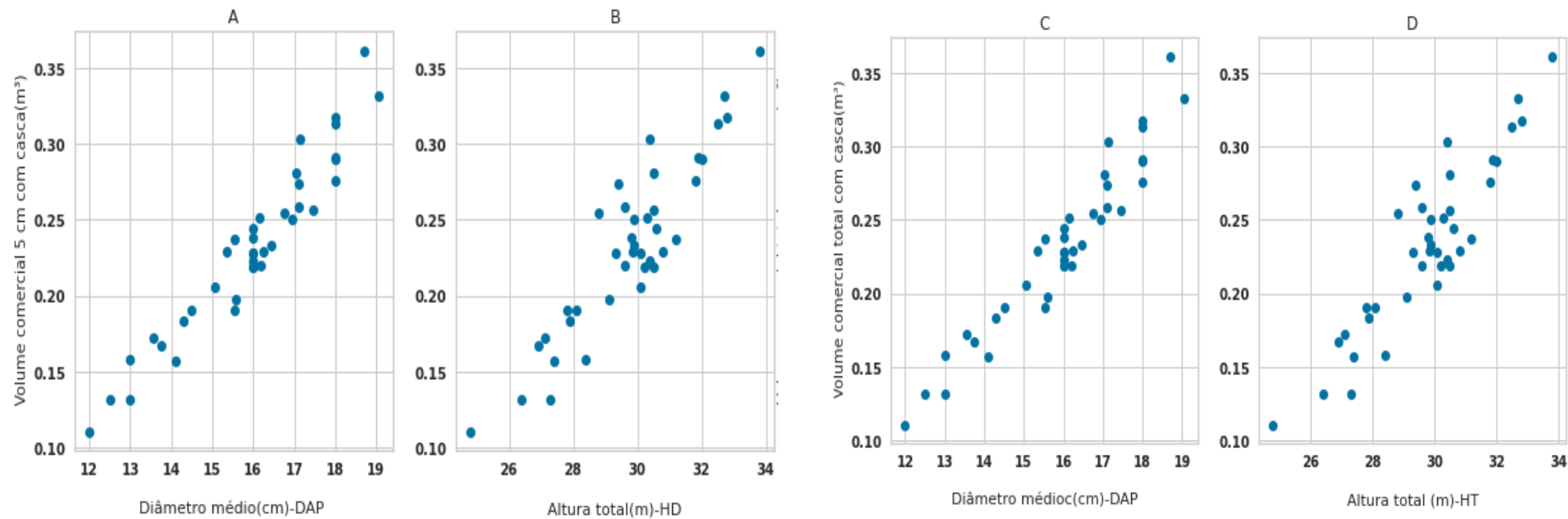
Smalian



V5cmcasca foi calculado no excel

Modelagem do volume:

Volumes das árvores totalizados (LnVt e LnVt5cm);



Modelagem do volume:

MODELOS AJUSTADOS (Schumacher e Hall):

- $\text{Ln}V_t = b_0 * \text{ln}d_{ap} + b_1 * \text{ln}h_t + e$

Coeficiente R^2 : 0.952277

Descrição do modelo: w: ['1.743337 ', '1.045491'] b: -9.85599

- $\text{Ln}V_{t5cm} = b_0 * \text{ln}d_{ap} + b_1 * \text{ln}h_t + e$

Coeficiente R^2 : 0.95

Descrição do modelo: w: ['1.743428 ', '1.045267 '] b: -9.85545

Erro de amostragem:

Amostragem Casual Simples

t95% 2.021075

| | V5casca | Vtcasca |
|--------------------|----------|----------|
| MÉDIA | 8.894588 | 8.895009 |
| VAR | 1.082255 | 1.082205 |
| DESVIO | 1.040315 | 1.040291 |
| CV | 11.69604 | 11.69522 |
| ERRO PADRÃO | 0.164488 | 0.164484 |

| | | |
|--------------------|----------|----------|
| ERRO DE AMOSTRAGEM | | |
| E% | 3.737589 | 3.737325 |

| | | |
|------------------------|----------|----------|
| INTERVALO DE CONFIANÇA | | |
| SUPERIOR | 9.227031 | 9.227445 |
| INFERIOR | 8.562145 | 8.562574 |

| | V5casca_ha | Vtcasca_ha |
|--------------------|-------------|-------------|
| MÉDIA | 222.3647016 | 222.3752323 |
| VAR | 676.4094392 | 676.3781057 |
| DESVIO | 26.00787264 | 26.00727025 |
| CV | 11.69604369 | 11.69521892 |
| ERRO PADRÃO | 4.112205732 | 4.112110485 |

| | | |
|--------------------|-------------|------------|
| ERRO DE AMOSTRAGEM | | |
| E% | 3.737588631 | 3.73732507 |

| | | |
|------------------------|-------------|-------------|
| INTERVALO DE CONFIANÇA | | |
| SUP | 230.6757794 | 230.6861176 |
| INF | 214.0536238 | 214.064347 |

Considerações finais:

- A Cubagem deveria ser feita em árvores no estrato de idade dos talhões do IPC (IPC 8 anos e Cubagem 2 anos);
- Foi utilizado os estimadores de amostragem aleatória simples considerando que o IPC foi para um estrato;
- Utilizou-se para a estimativa dos volumes uma abordagem paramétrica, a qual permite analisar o impacto do parâmetro ajustado de cada variável explicativa no valor do volume estimado;
- Para estimação da altura utilizou-se modelos abordagem não paramétrica. Além de permitir considerar variáveis categóricas , foi possível melhorar as estimativas em relação ao ajuste de modelos regressivos.