1) Classificação - Árvores de decisão

A.

Código: [github](https://github.com/SalvadorBrunoo/DataMining/blob/main/EX3/dataset/createDf.ipynb)

Resultado: [soybean.csv](https://github.com/SalvadorBrunoo/DataMining/blob/main/EX3/dataset/soybean.csv)

B. Código: [github](https://github.com/SalvadorBrunoo/DataMining/blob/main/EX3/ex3.1.ows)

Muitas features atrapalharam na investigação de relações intuitivas, até mesmo a visualização não traz interpretação fácil dos dados.

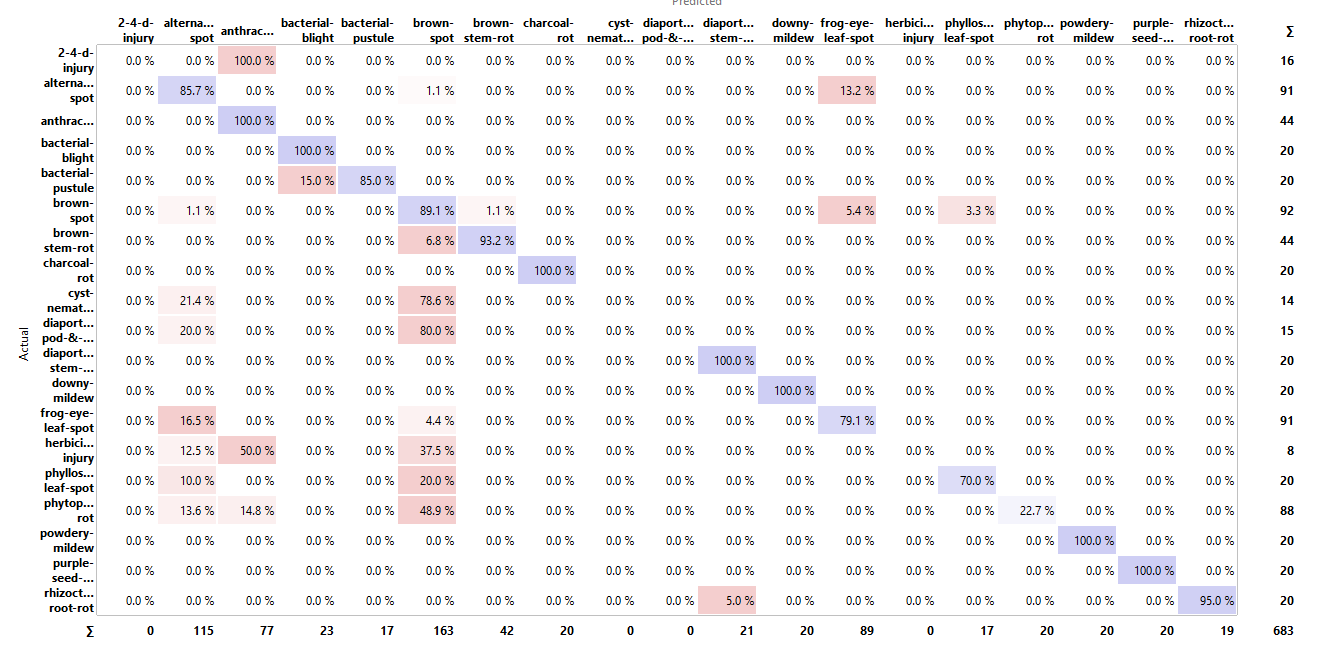
C. Código: [github](https://github.com/SalvadorBrunoo/DataMining/blob/main/EX3/ex3.1.ows)

87 NÓS, 44 FOLHAS



F1 considera ao mesmo tempo o acerto em relação aos falsos positivos e o acerto em relação aos falsos negativos, que são o *Precision* e *Recall* respectivamente. Já a medida CA (*Accuracy*), não considera *Precision* e *Recall*, apenas os verdadeiros positivos. A medida AUC (*Area Under Curve*) demonstra a capacidade de acerto dos verdadeiros positivos e também dos verdadeiros negativos, porém sem levar em conta *Precision e Recall.* Como queremos além de saber o acerto poder mensurar os casos de falsos positivos e negativos de maneira equilibrada, então F1 é uma mais completa para a análise do modelo.

D. Código: [github](https://github.com/SalvadorBrunoo/DataMining/blob/main/EX3/ex3.1.ows)

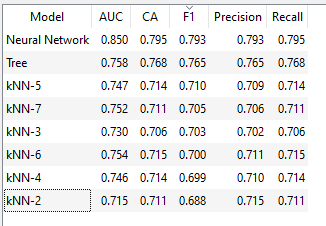


A classe 2-4-d-injury, praticamente não tem dados preenchidos, que determina 0,0% de acerto. O mesmo acontece para as classes: cyst-nematode, diaporthe-pod-&-stem-blight, herbicide-injury.

Algumas classes apresentam 100 % de acerto, possivelmente porque o dataset contém features muito significativas para a inferência dessas classes especificas, já que mesmo com pouca quantidade em relação as demais, o treinamento do modelo foi efetivo.

2) Classificação

Código: [github](https://github.com/SalvadorBrunoo/DataMining/blob/main/EX3/ex3.2.ows)

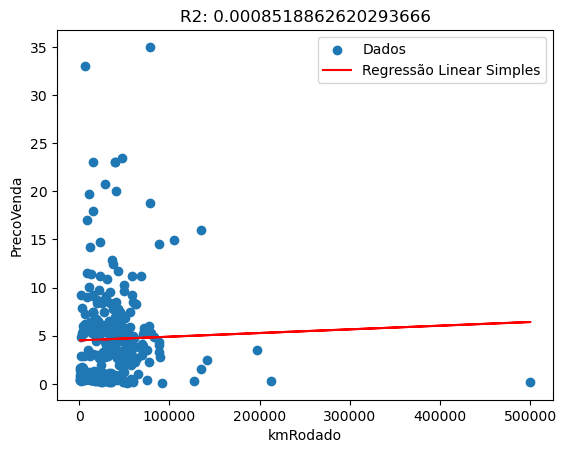


2.1, 2.2, 2.3 e 2.4)

3) Regressão

Código: [github](https://github.com/SalvadorBrunoo/DataMining/blob/main/EX3/ex3.3.ipynb)

3.1)



3.2)

Coeficientes:

[ 3.82457823e-01 -2.86702234e-04 1.64623728e-06 5.58625742e+00 4.73812925e+00 8.51402501e-01]

Intercept:

-777.630241319014

R² nos dados de teste:

0.639177323941603

Erro Quadrático Médio nos dados de teste:

8.378292953723644