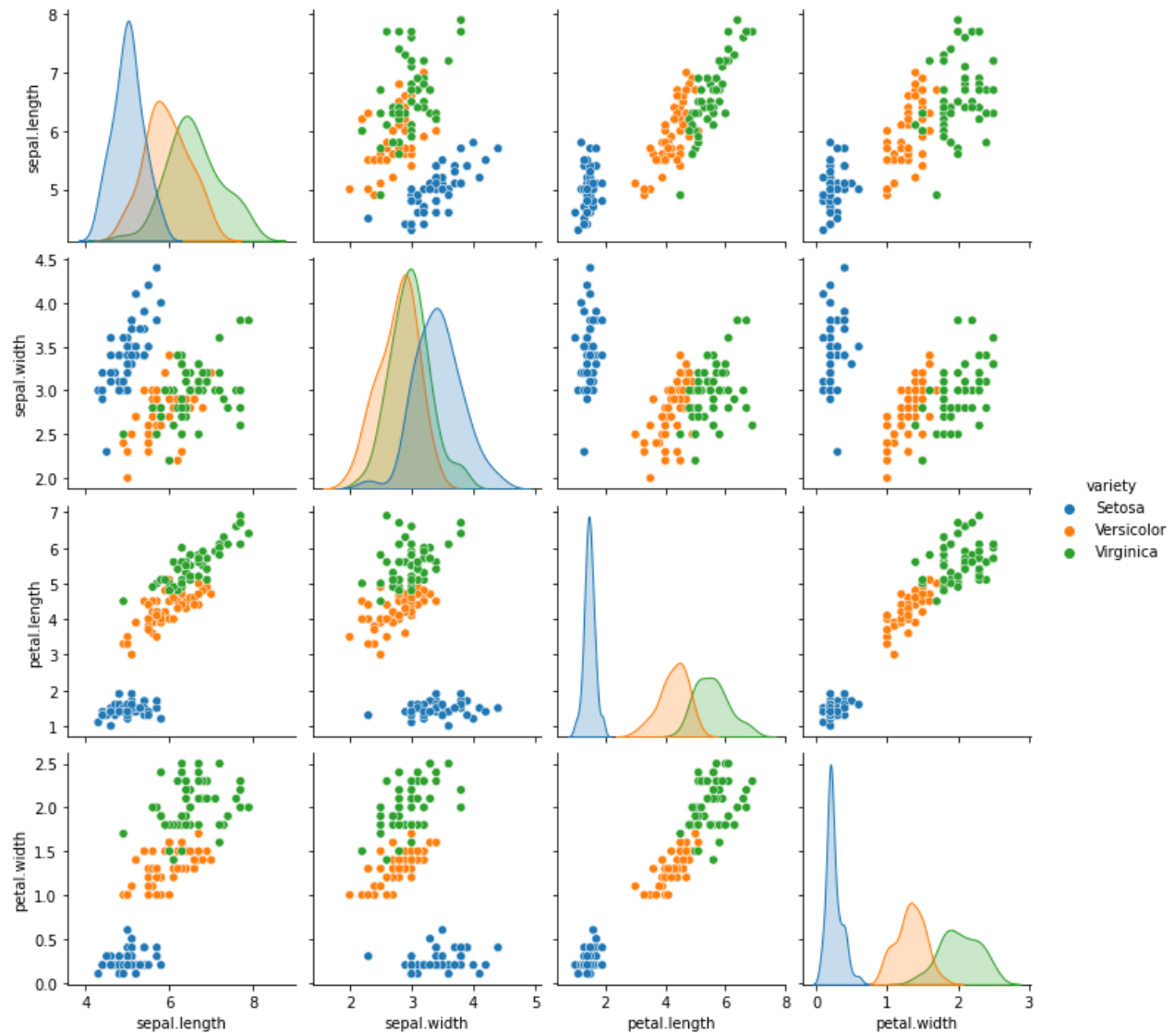


Árvore de Decisão

Utilizando a base de dados Iris, com 70% em treinamento e 30% em teste, foi aplicada uma implementação de árvore de decisão.

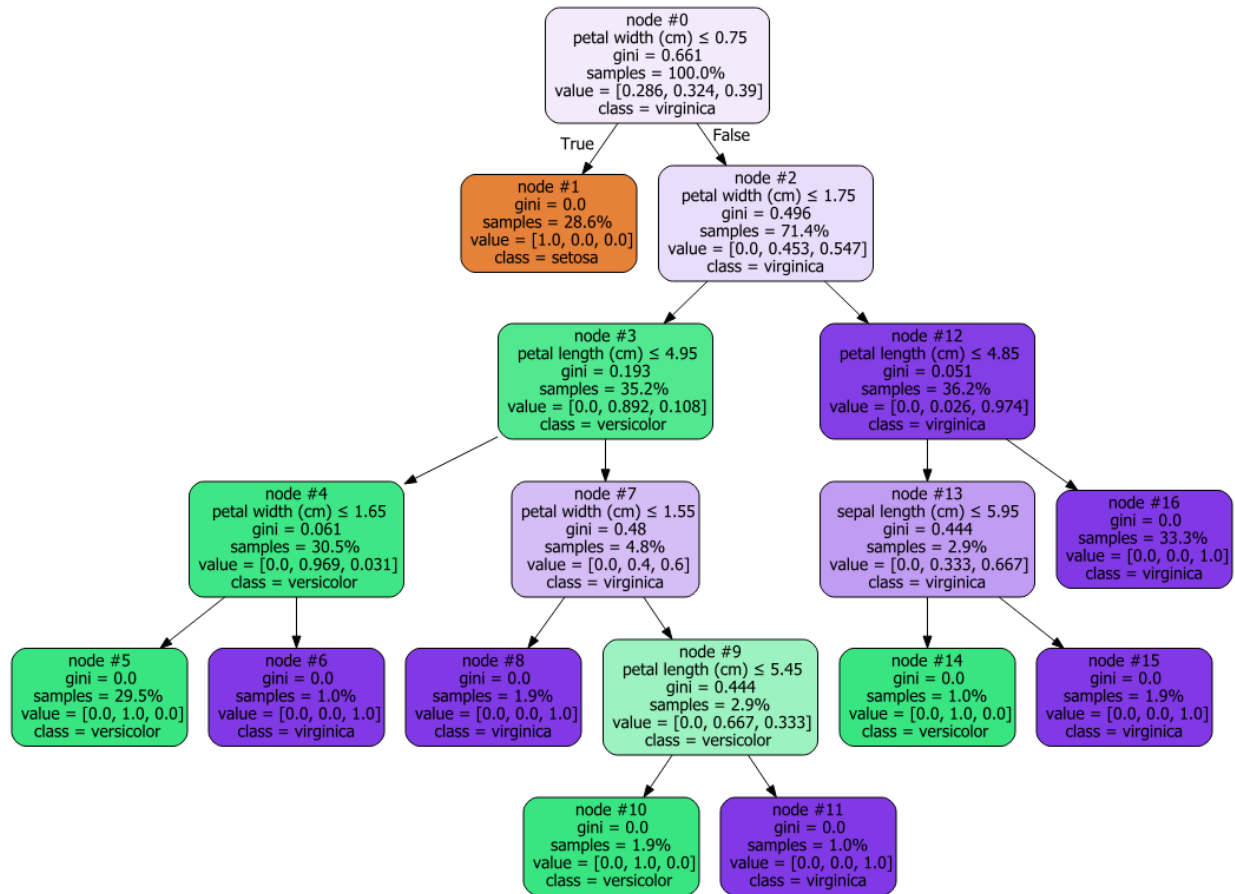
Segue uma plotagem dos pares de atributos da base de dados:



Árvores de decisão foram geradas levando em consideração esses atributos, e em geral, o tamanho da árvore variou entre 5 e 19 nós, com cerca de 10 nós em média. A acurácia, por sua vez, variou entre 85% e 100%.

Dois casos em especial serviram bem como demonstração dos extremos:

Caso 1

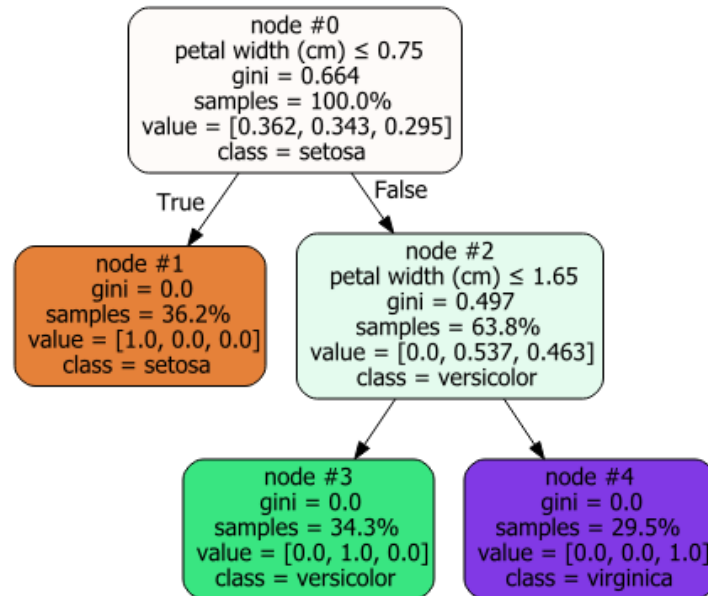


Com **17 nós**, esta árvore obteve **100% de acurácia** tanto no conjunto de treinamento quanto de testes, ignorando o atributo *sepal width* por completo, e usando o *sepal length* para discernir somente 2.9% das amostras.

Segue a tabela de importância de atributos desta árvore:

Atributo	Importância
Sepal Length	0.01920
Sepal Width	0.0
Petal Length	0.06836
Petal Width	0.91242

Caso 2



Esta segunda árvore teve somente **5 nós**, mas obteve uma **acurácia de 88%** no conjunto de testes, usando somente *petal width* como atributo de definição.

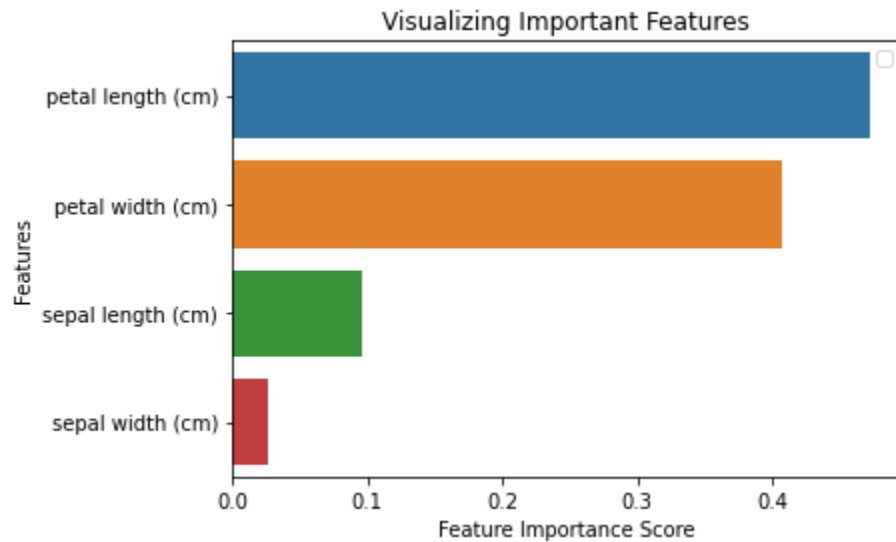
Sua tabela de importância de atributos foi igualmente simples:

Atributo	Importância
Sepal Length	0.0
Sepal Width	0.0
Petal Length	0.0
Petal Width	1.0

Random Forest

Novamente utilizando a base de dados Iris, com 70% em treinamento e 30% em teste, foi aplicada uma implementação do Random Forest.

De forma randômica, a Random Forest, com 500 árvores de decisão a cada iteração, conseguiu em média 95.4% de acurácia no conjunto de testes, com a seguinte distribuição de importância de atributos:



Foram feitos testes com 100 iterações para conseguir uma média para variações na quantidade de árvores de decisão, tendo 500 árvores como melhor resultado em geral:

Árvores	Média de Acurácia
100	95,3%
200	95,1%
300	95%
400	95,3%
500	95,4%
600	95,4%
700	95,2%
800	95,3%
900	94,4%
1000	94,4%

Além disso, foi experimentado a remoção de diferentes atributos, seguindo o mesmo procedimento de 100 iterações para obtenção das médias de acurácia:

Sepal Length	Sepal Width	Petal Length	Petal Width	Acurácia
o	o	o	o	95,2%
	o	o	o	95,2%
o		o	o	94,4%
o	o		o	93,3%
o	o	o		93,3%
		o	o	95,8%
o			o	92,7%
o	o			71,0%
	o	o		93,3%
o		o		93,3%
	o		o	93,0%
o				68,0%
	o			50,1%
		o		93,0%
			o	95,2%

Resultados relevantes são o consideravelmente alto grau de acurácia utilizando somente o *Petal Width*, obtendo praticamente o mesmo resultado de 95,2% que quando se utilizou todos os atributos.

Assim como o grau mais alto de acurácia obtido utilizando somente *Petal Length* e *Petal Width*, obtendo 95,8% de acurácia, sendo melhor até do que quando todos atributos foram considerados.

Em contrapartida, o menor grau de acurácia ocorreu quando se utilizou somente o *Sepal Length*, com 50,1% de acurácia, seguido de somente o *Sepal Width*, com 68,0% de acurácia.