

# B2W-Challenge

...

Angelo Mendes

# Sumário

- Introdução
- Tratamento dos dados
- Análise de Correlação
- Árvore de Decisão
- Regras de Associação
- Conclusões

# Introdução

- Projeto desenvolvido em Python3.7 com bibliotecas: pandas, matplotlib, numpy, sklearn e apyori
- O versionamento e disponibilizado do código fonte foi feito com Git e Github. O link do repositório é: <https://github.com/AngeloMendes/B2W-Challenge>
- Cada tarefa do sumário foi organizada em um issue que possui as análises e conclusões
- Cada tarefa possui sua branch e um pull request foi realizado na branch develop, e na conclusão de todo o desafio foi feito o merge da develop na master

# Tratamento dos Dados

Analisar os dois datasets e verificar a necessidade dos tratamentos abaixo:

1. Verificar a existência de dados faltantes
2. Normalização dos dados
3. Discretização dos dados
4. Tratamento de outliers

# Tratamento dos Dados

Link para acesso: <https://github.com/AngeloMendes/B2W-Challenge/issues/1>

1. Não há dados 'nulos' ou 'nan'
2. Não há necessidade de normalizar por os valores numéricos são apenas os valores das compras contendo 2 casas decimais. Normalizar acarretaria em números com mais casas decimais e poderia atrapalhar nos cálculos, ideia oposta à aplicação da normalização.

# Tratamento dos Dados

3. A discretização não foi necessária, pois atributo categórico, PAY\_TYPE, já estava discretizado. Os demais não há necessidade de discretizar nesta tarefa.

4. Foi identificado alguns outliers no atributo COMPETITOR\_PRICE, como o valor 148000.00 no produto P2 para o competidor C2 na data 2015-10-14 08:11:39. A correção dos valores neste e nos demais casos foi feita com a substituição do valor incorreto pelo valor anterior que não destoava dos demais. No exemplo citado anteriormente, analisou uma compra feita em 2015-10-14 08:10:28 com valor de 1480.00 para o mesmo produto e mesmo competidor, logo este valor substituiu o valor outlier.

# Tratamento dos Dados

4. Ainda no tratamento dos outliers, posteriormente viu-se que havia uma recorrência considerável dos erros, então desenvolveu-se um método para ranqueamento dos maiores e menores valores de todos os PROD\_ID agrupados por COMPETITOR, independentemente do tipo de pagamento, para que pudesse ser avaliado aos valores corretos. Um exemplo de registro incorreto:

PROD_ID	DATE_EXTRACTION	COMPETITOR	COMPETITOR_PRICE	PAY_TYPE
P5	2015-10-10 08:11:33	C2	853.52	1
P5	2015-10-14 08:11:39	C2	84890.0	1

# Tratamento dos Dados

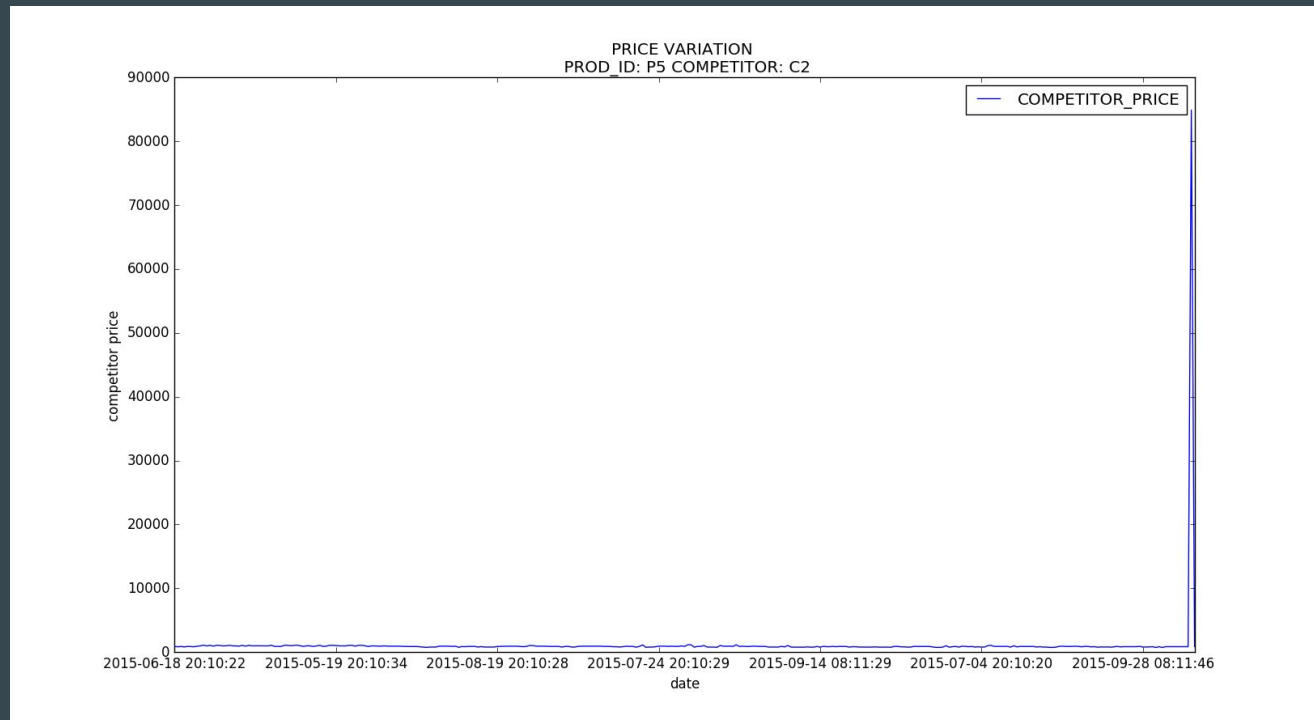
Além da correção pelo valor da última venda, viu-se cenários em que o erro era devido a pontuação das casas decimais, então fez-se apenas esta correção. Os valores corretos ficaram assim:

PROD_ID	DATE_EXTRACTION	COMPETITOR	COMPETITOR_PRICE	PAY_TYPE
P5	2015-10-10 08:11:33	C2	853.52	1
P5	2015-10-14 08:11:39	C2	848.90	1



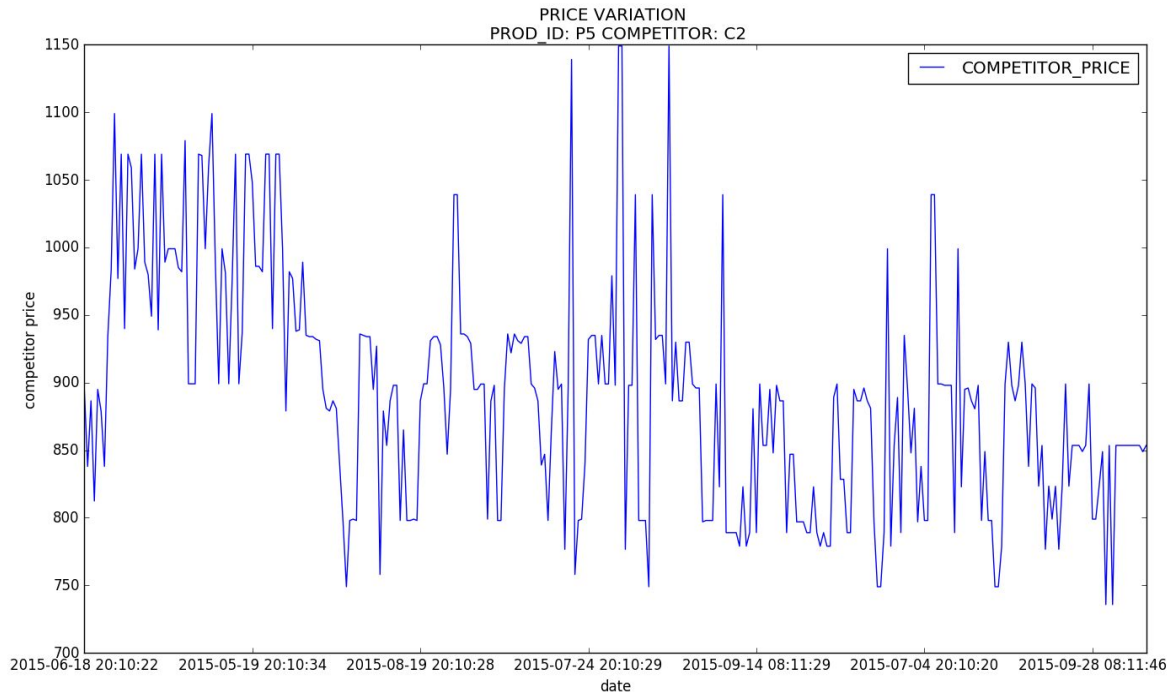
# Tratamento dos Dados

Imagem com outlier



# Tratamento dos Dados

Imagem sem outlier



# Análise de Correlação

Link para acesso: <https://github.com/AngeloMendes/B2W-Challenge/issues/2>

Para fazer a correlação foi necessário discretizar os atributos PROD\_ID e COMPETITOR, além de converter o atributo DATE\_EXTRACTION de timestamp para inteiro.

# Análise de Correlação

O resultado é mostrado a seguir:

	PROD_ID	DATE_EXTRACTION	COMPETITOR	COMPETITOR_PRICE	PAY_TYPE
PROD_ID	1.000000	0.121392	0.005499	-0.279069	-0.000009
DATE_EXTRACTION	0.121392	1.000000	-0.048366	-0.301933	-0.000282
COMPETITOR	0.005499	-0.048366	1.000000	-0.065275	-0.000006
COMPETITOR_PRICE	-0.279069	-0.301933	-0.065275	1.000000	-0.024200
PAY_TYPE	-0.000009	-0.000282	-0.000006	-0.024200	1.000000

# Análise de Correlação

Como conclusão sobre as correlações, pode-se considerar:

- a data da compra tem maior peso na compra de um determinado produto
- o vendedor/marca também influencia no produto
- tipo de pagamento não interfere tanto na escolha de quem está vendendo o produto

Ações de marketing com intuito de fidelização do cliente apresentam-se como diferencial neste cenário.

# Árvore de Decisão

Link para acesso: <https://github.com/AngeloMendes/B2W-Challenge/issues/4>

Gerou-se a árvore de decisão de cada produto a fim de avaliar a precisão (índice gini) da predição do COMPETITOR. O atributo mais relevante à classificação é o COMPETITOR\_PRICE, seguido por DATE\_EXTRACTION e PAY\_TYPE com bem menos relevância.

Fez-se uma poda por meio de limitação do número de nós devido a limitação computacional, mas o índice gini não reduziu tanto com o aumento dos nós da árvore, o que nos permite deduzir que há mais fatores impactando na compra dos produtos.

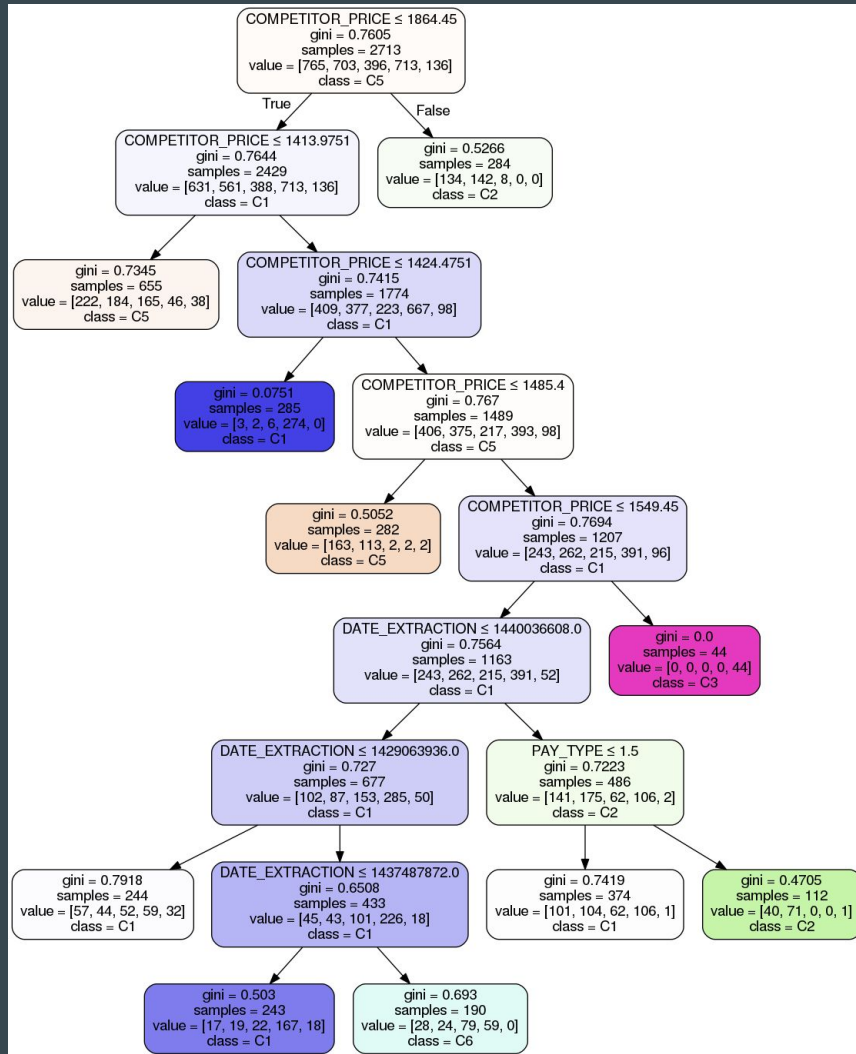
# Árvore de Decisão

Como não há um rótulo indicando a conclusão da compra no preço oferecido, não há como avaliar uma combinação de atributos neste cenário.

Tentou-se cruzar as informações de preço dos dois datasets, mas não fez muito sentido pois os preços no dataset 'sales.csv' possui valores com variações, como descontos, e a data não está como timestamp.

# Árvore de Decisão

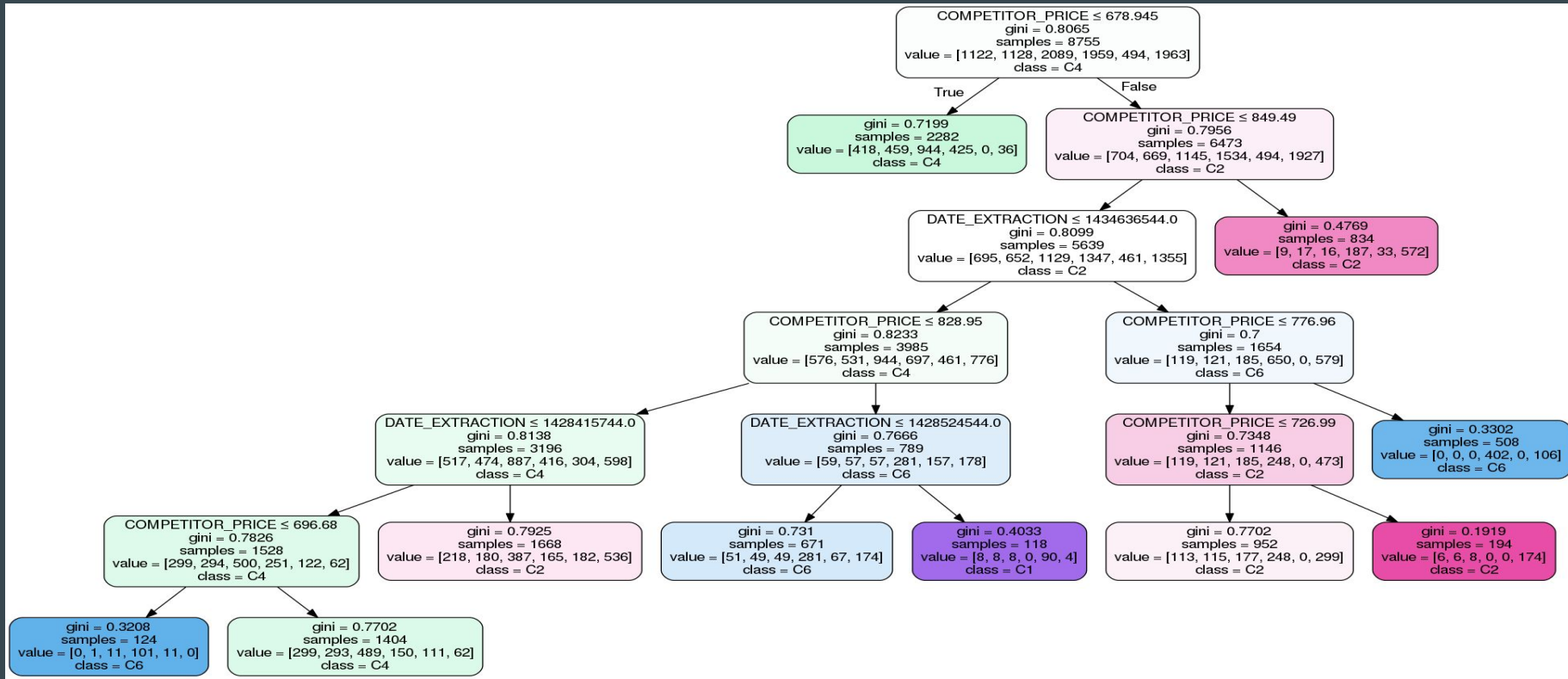
## Árvore do Produto P1





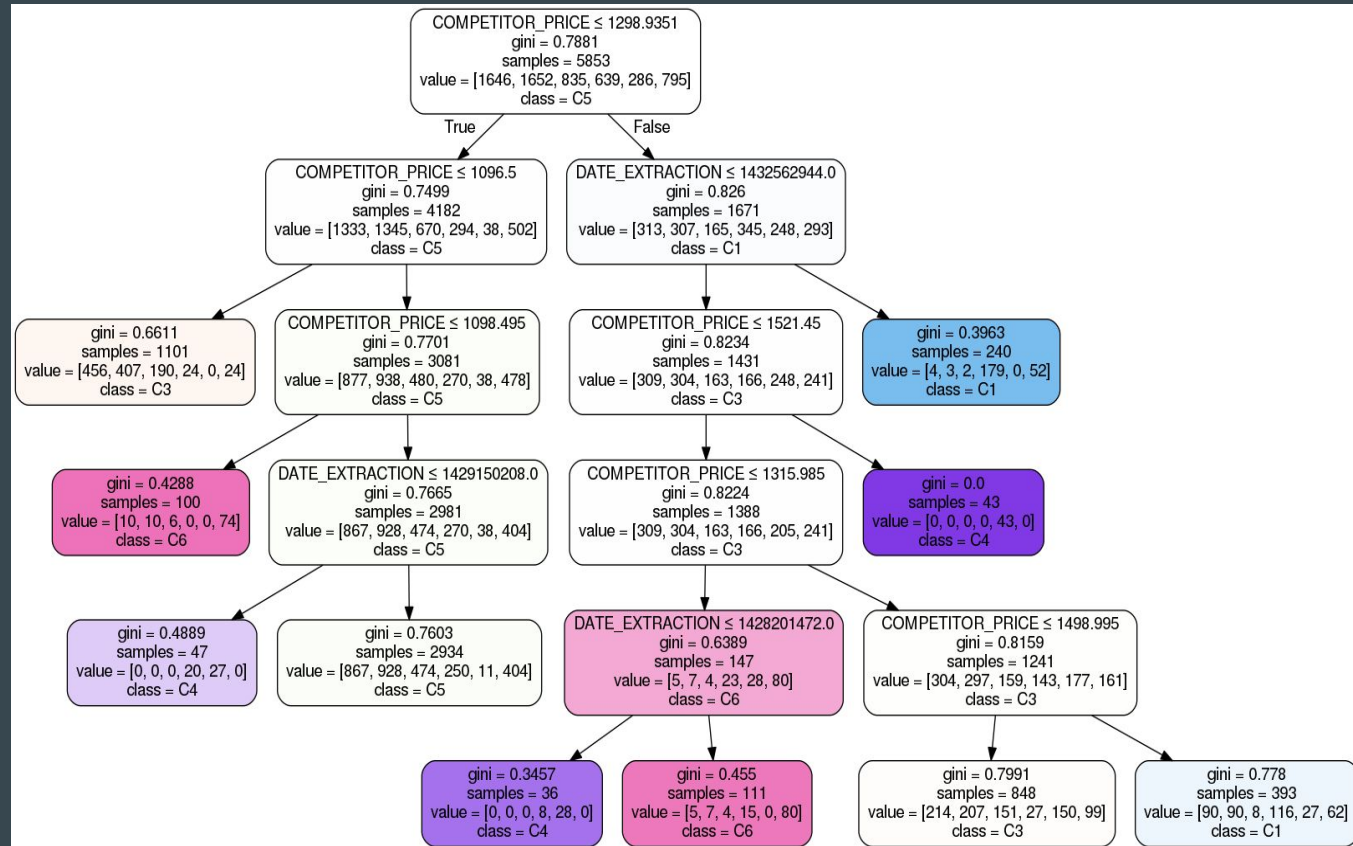
# Árvore de Decisão

## Árvore do Produto P2



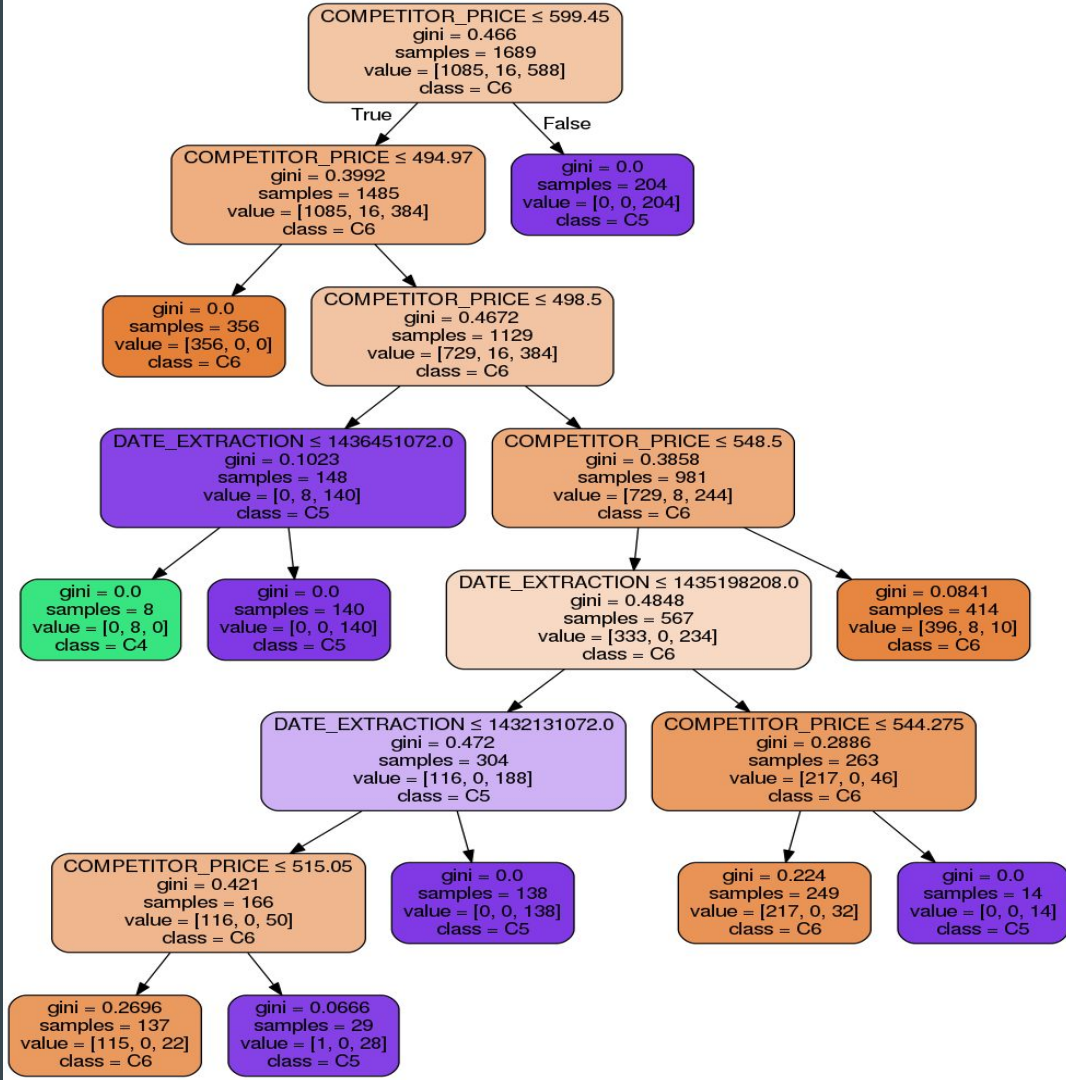
# Árvore de Decisão

## Árvore do Produto P3



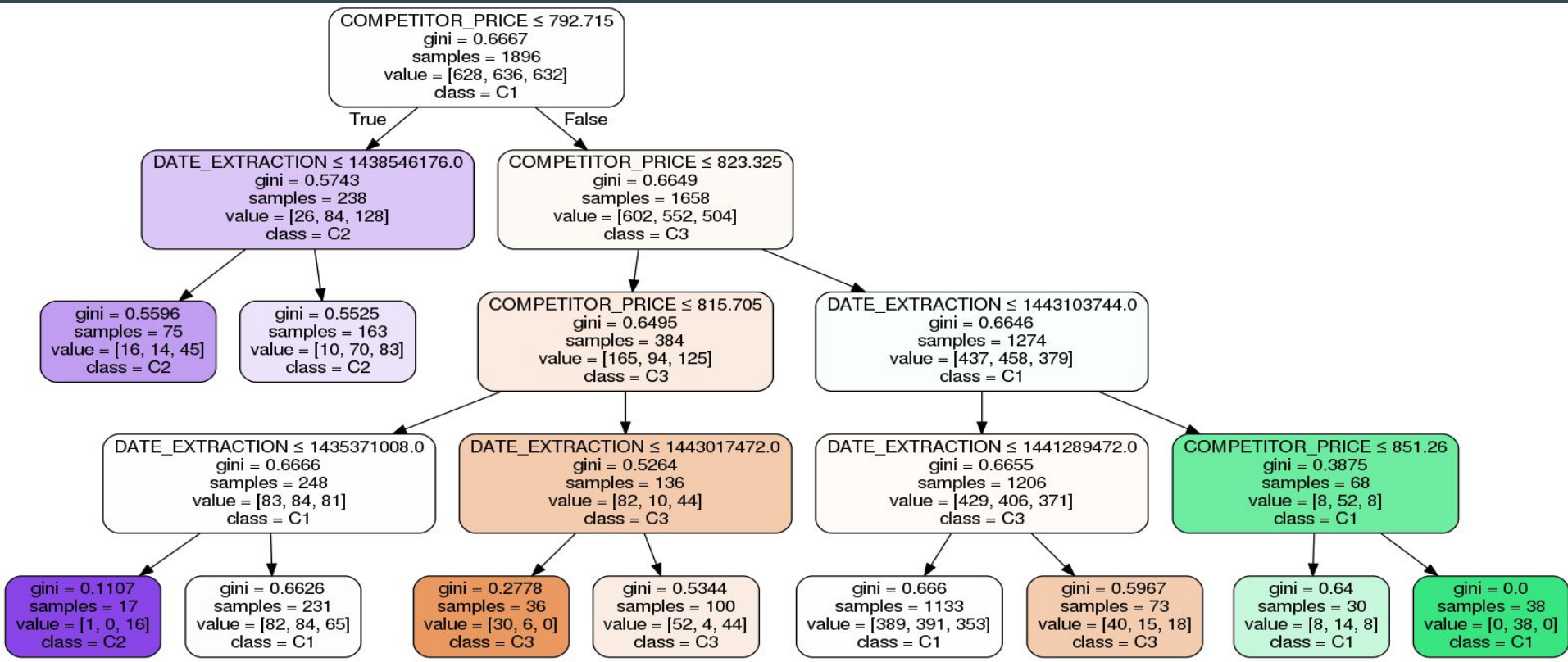
# Árvore de Decisão

## Árvore do Produto P4



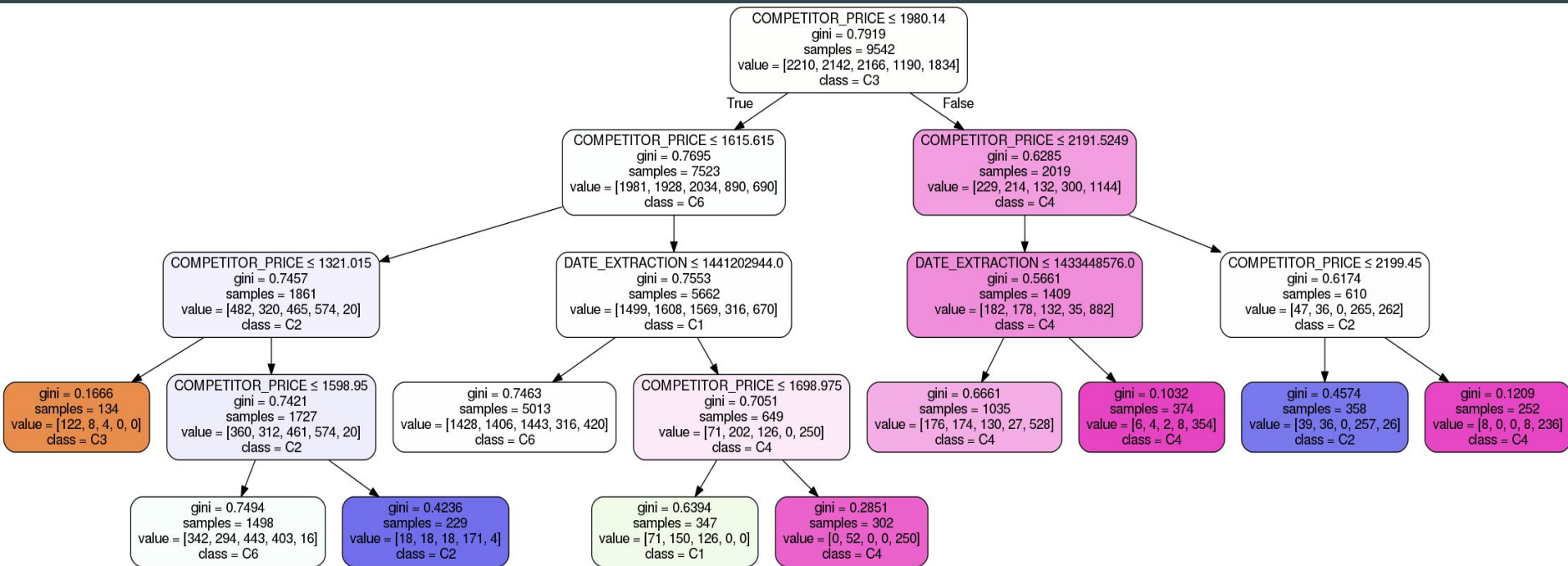
# Árvore de Decisão

## Árvore do Produto P5



# Árvore de Decisão

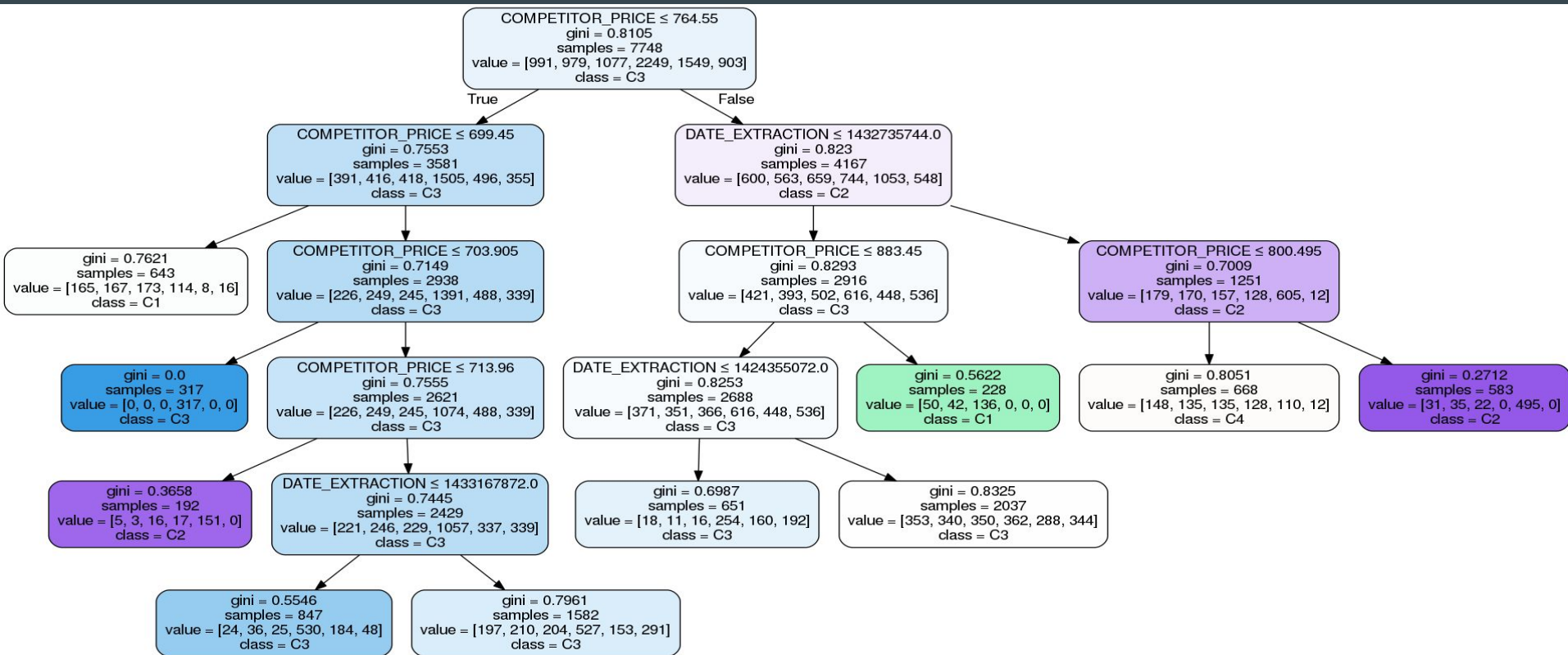
## Árvore do Produto P6





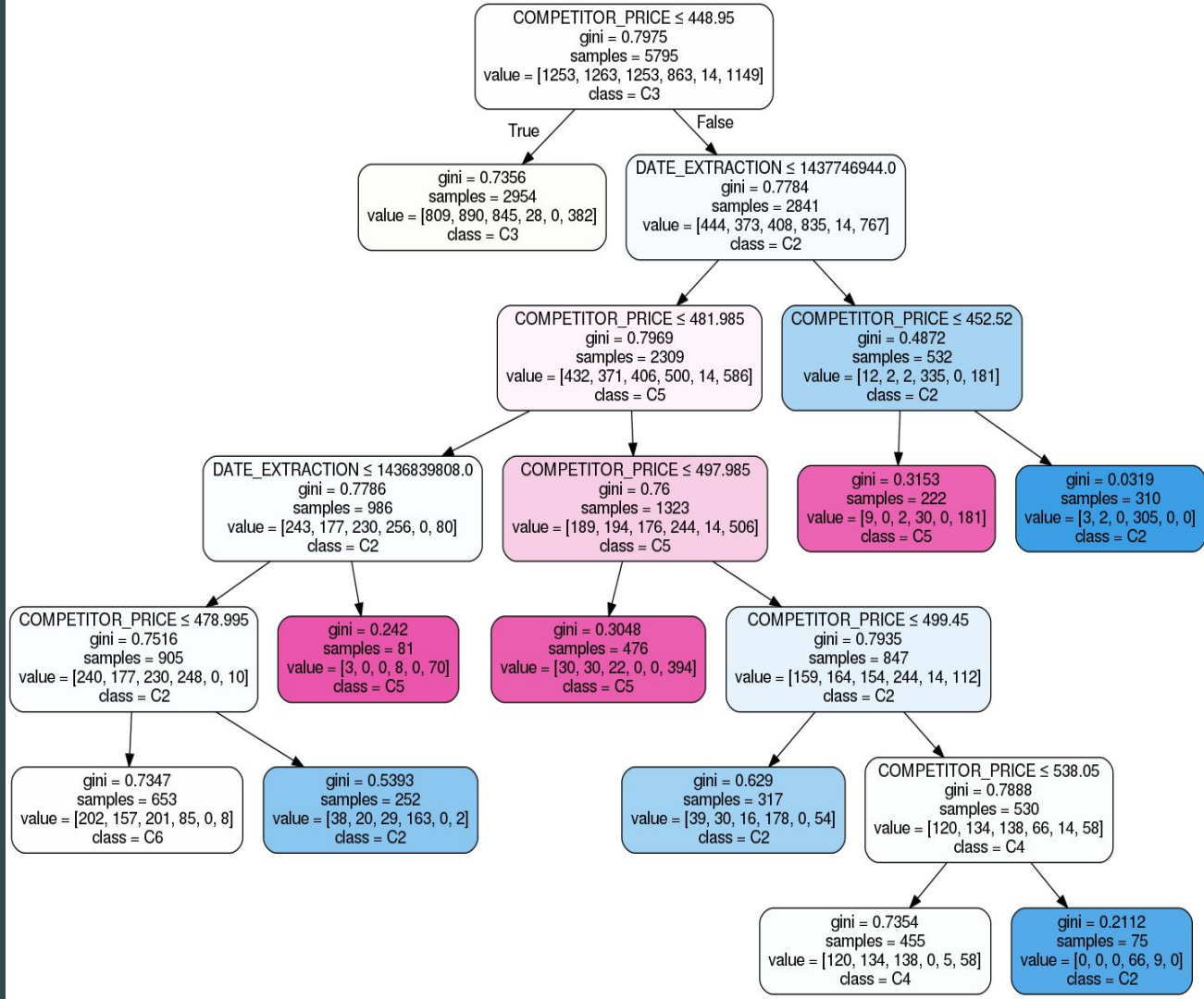
# Árvore de Decisão

## Árvore do Produto P7



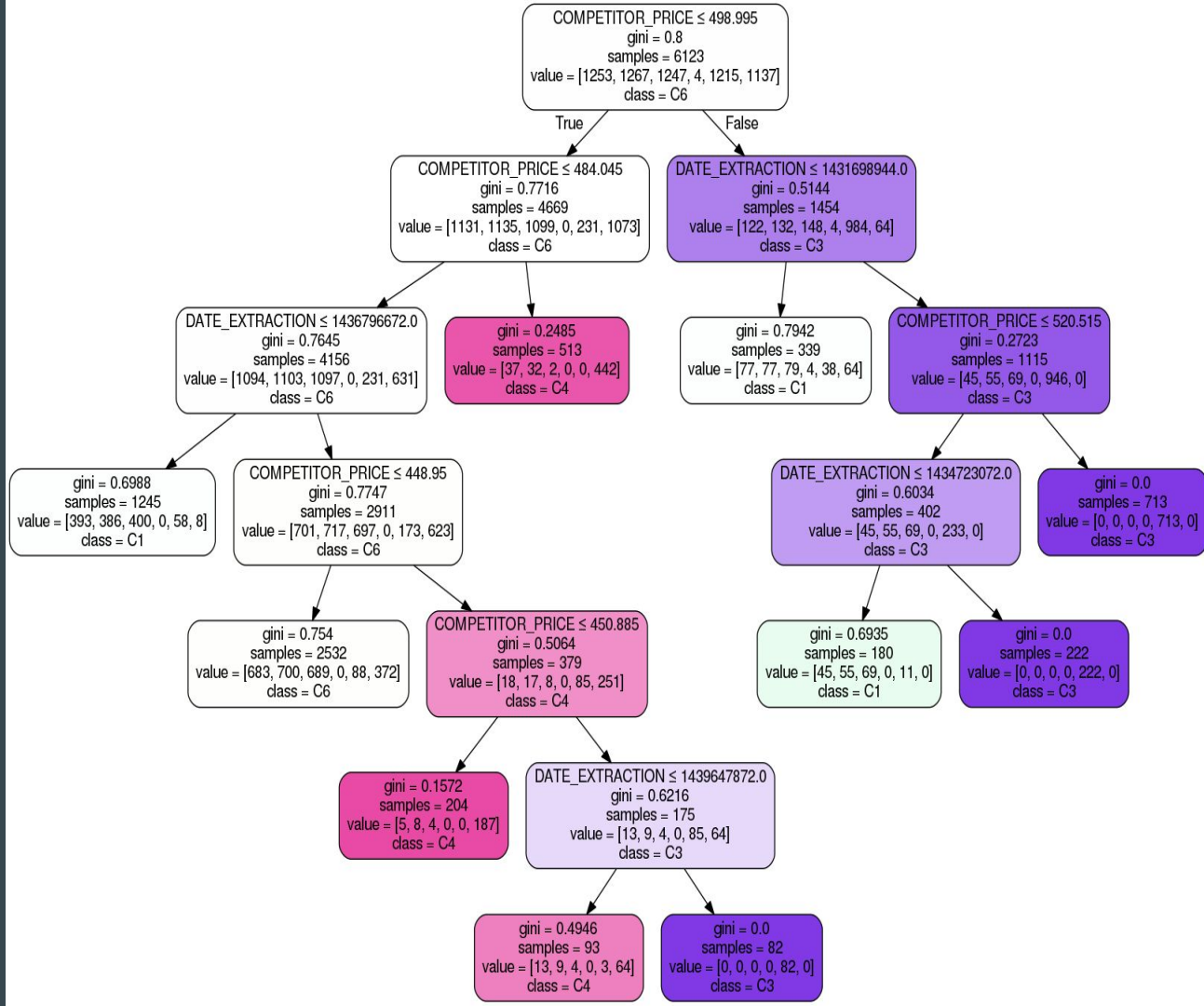
# Árvore de Decisão

## Árvore do Produto P8



# Árvore de Decisão

## Árvore do Produto P9





# Regras de Associação

Link para acesso: <https://github.com/AngeloMendes/B2W-Challenge/issues/5>

Devido à disposição das regras, será apresentado apenas as conclusões de cada produto, todas podem ser acessadas nos links acima.

As regras seguem a seguinte estrutura:

=====

Rule: 2015-08-10 20:10:18 -> C2

Support: 0.01

Confidence: 1.0

Lift: 11.111111111

=====

# Regras de Associação

No produto P1 destaca-se que:

- o preço entre 1403.9 e 1424.05 pode ser considerado bom ao produto visto que foi pago imediatamente
- as duas últimas regras com confiança 1 mostram as vendas no mesmo dia para C1 e C2. Eles venderam com o mesmo preço e o mesmo tipo de pagamento

PROD\_ID,DATE\_EXTRACTION,COMPETITOR,COMPETITOR\_PRICE,PAY\_TYPE

P1, 2015-08-10 20:10:18, C1, 1425.9, 2

P1, 2015-08-10 20:10:18, C2, 1425.9, 2

# Regras de Associação

No produto P2 destaca-se que:

- 788.0 é o melhor preço
- 829.0 é um preço elevado, devido ao tipo de pagamento utilizado para compras neste valor
- C6 fez uma venda por 879.0 com pagamento imediato, suas estratégias para relacionamento com o cliente fez a diferença
- a data também apresenta relação com a compra, mesmo com a confiança um pouco inferior às demais regras

# Regras de Associação

No produto P3, destaca-se:

- o preço entre 1274.15 e 1304.13 é o mais indicado
- 1319.0 já é elevado, visto que a compra feita em 2015-04-11 08:11:25 não teve pagamento imediato
- C5 conseguiu fazer vendas com preço igual à 1274.15, 1304.13 e 1543.9. Sendo está última sem pagamento imediato

# Regras de Associação

No produto P4, destaca-se:

- 499.0 é o melhor preço, porém 519.0 pode ser aceitável visto que há uma boa confiança nas vendas com esse valor em ambos tipos de pagamento
- C6 possui os melhores preços

# Regras de Associação

No produto P5, destaca-se:

- a data foi extremamente importante na definição do preço e na quantidade de vendas

# Regras de Associação

No produto P6, destaca-se:

- 1869.15 e 1999.0 é a faixa de melhores preços
- C2 mostrou-se ofertar o melhor preço
- C6 conseguiu vendas até com preço alto no valor de 2379.87, mesmo sem pagamento imediato

# Regras de Associação

No produto P7, destaca-se:

- 721.65 é o melhor preço, porém C1 e C3 conseguiram vendas com pagamento imediato no valor de 899.0
- 2015-02-12 20:10:24 não é um bom dia para venda do produto, devido ao valor e como foi feito os pagamento nesta data



# Regras de Associação

No produto P8, destaca-se:

- o preço ideal está entre 449.9 (aplicado por C6) e 459.9
- C4 realizou venda com pagamento imediato por 449.9, já com o preço de 479.0 e não teve mais o mesmo tipo pagamento

# Regras de Associação

No produto P9, destaca-se:

- 443.0 é o menor preço, porém C6 conseguiu vendas por 449.9
- C5 realizou vendas por 599.0, porém é um valor alto e, por isso, não foi feito pagamento imediato
- 2015-08-10 20:10:18 foi a data com menor preço

# Conclusões

Com as análises pode-se inferir algumas informações sobre a precificação do produto, como melhor data para venda, o tipo de pagamento serve como indicador para preço elevado, avaliar como concorrentes vendem com preços mais altos são as principais

Informações sobre tipo do produto e localização das concorrentes podem ser úteis para relacionar com a campanha de precificação que é feita em datas específicas

Avaliar a precificação apenas com o número de vendas pode acarretar em uma avaliação não muito eficiente