



PROPOSTA ORIENTADA A DADOS PARA MAXIMIZAR LUCROS DE CAMPANHA

ROADMAP





PROBLEMA

- ✗ Empresa prevê dificuldades financeiras nos próximos 3 anos
- ✗ Quer vender um produto e fez uma propaganda teste com ele mas deu prejuízo
- ✗ Vender o produto para obter o maior lucro possível
- ✗ Contratar equipe de analistas de dados para desenvolver solução
- ✗ Explicar para o CMO sobre as características do consumidor deste produto



MÉTODOS UTILIZADOS

1. Tratamento dos dados
2. Análises exploratórias
3. Visualizações
4. Segmentações dos clientes
5. Criação de modelos de aprendizagem de máquina

CONVENÇÕES

1. Clientes que adquiriram o produto teste serão nomeados como clientes positivos
2. Clientes que não adquiriram serão nomeados como clientes negativos

PRINCIPAIS RESULTADOS

Classificação do cliente	Quantidade
Negativo	1906
Positivo	334

PRINCIPAIS RESULTADOS

Características que quanto maiores, **menor** a chance do cliente ser positivo

	Recency	Dt_Customer	Teenhome	Kidhome	Married
Negativo	51,50	26/07/2013	0,54	0,46	0,40
Positivo	35,38	07/04/2013	0,30	0,34	0,14

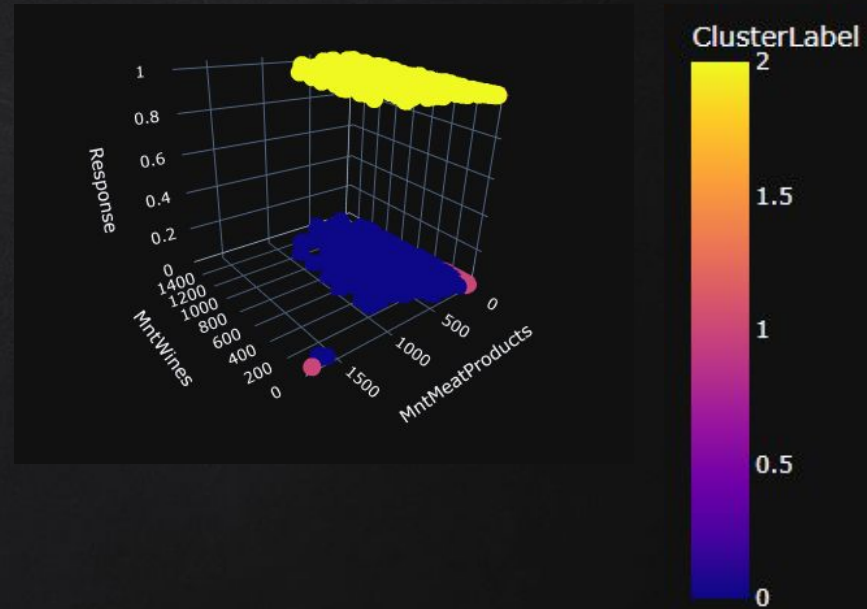
PRINCIPAIS RESULTADOS

Características que quanto maiores, **maior** a chance do cliente ser positivo

	AcceptedC5	AcceptedC1	AcceptedC3	MntWines	MntMeat
Negativo	0,04	0,03	0,05	269,10	144,62
Positivo	0,27	0,24	0,23	502,7	294,35

PRINCIPAIS RESULTADOS

Conjunto	MntWines	MntMeat Products	Classificação
0	386	156	Negativo
1	15	11	Negativo
2	460,5	203	Positivo



PRINCIPAIS RESULTADOS

Metodologia 1

10-Fold Cross validation

LR: Mean Accuracy=0.89396, Standard Deviation=0.00945
KNN: Mean Accuracy=0.87723, Standard Deviation=0.02272
CART: Mean Accuracy=0.84206, Standard Deviation=0.01187
SVM: Mean Accuracy=0.87501, Standard Deviation=0.01113
NB: Mean Accuracy=nan, Standard Deviation=nan

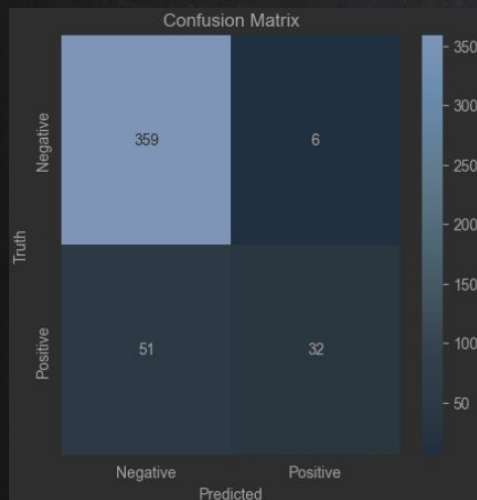
Metodologia 2

10-Fold Cross validation

LR: Mean Accuracy=0.89396, Standard Deviation=0.00945
KNN: Mean Accuracy=0.87723, Standard Deviation=0.02272
CART: Mean Accuracy=0.83480, Standard Deviation=0.01730
SVM: Mean Accuracy=0.87501, Standard Deviation=0.01113
NB: Mean Accuracy=0.87611, Standard Deviation=0.01475

PRINCIPAIS RESULTADOS

Metodologia 1
(todas características)



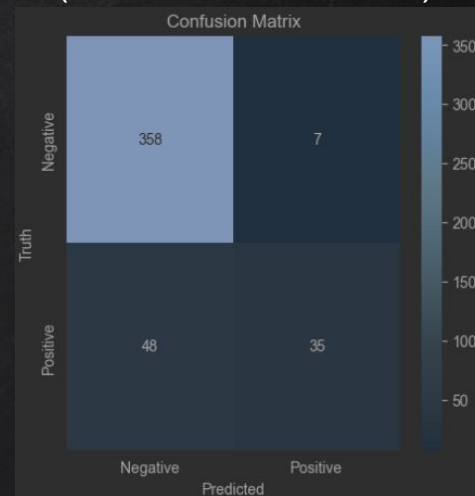
Cost-Revenue
Confusion Matrix

3	-3
-8	8

	0	1
Recall	98%	39%
Accuracy	87%	87%

Lucro de 907 MUs

Metodologia 2
(mais correlacionadas)



	0	1
Recall	98%	56%
Accuracy	88%	88%

Lucro de 949 MUs

COMENTÁRIOS FINAIS DOS RESULTADOS

- ✗ Independente das duas metodologias, os modelos ficaram melhores em prever se um cliente não vai comprar do que se ele vai comprar, pelo fato da base de dados estar desbalanceada
- ✗ Era esperado que a metodologia 1 fosse melhor que a 2, por apenas características mais correlacionadas ao rótulo
- ✗ Regressão logística foi o melhor modelo

SOLUÇÃO

- ✗ Utilizar a metodologia 2 para ter o maior lucro possível (949 MUs), 3995 MUs a mais que a campanha teste
- ✗ Se o cliente tiver a pouco tempo sem comprar algo, ser cliente de longa data, não ter crianças ou adolescentes, não ser casado, ter aceitado as campanhas anteriores 1, 3 ou 5 e gastou muito com vinhos e carne, a chance do cliente ser positivo é muito alta!

TRABALHOS FUTUROS

- ✕ Experimentar usar redes neurais artificiais em vez de apenas modelos tradicionais
- ✕ Experimentar usar processamento de linguagem natural em redes sociais para saber o que as pessoas acham de produtos similares ao que a empresa está propondo

FIM



OBRIGADO!