

Fármacia: Gestão de vendas

Presentation Report



AIAD 4MIEIC02 GRUPO 36

João Guedes | up201303880

Nuno Pereira | up201506265

Simão Lúcio | up201303485



Descrição do problema de análise de dados

Nesta fase do projeto, o principal objetivo é efetuar uma previsão, baseada em dados de transações já realizadas, de forma a perceber quais os medicamentos mais procurados pelos clientes e por que preço poderão ser vendidos. Com base nessa previsão, o agente Farmácia terá de ser capaz de **estabelecer preços** que **potenciem o lucro da farmácia**, mas que de facto se convertam em compras dos Clientes.

Como tal, para isso, iremos precisar de iniciar vários agentes farmácia e cliente, que efetuem múltiplas transações, para que possamos obter uma boa base de dados capaz de obter informação relevante para futuramente ser utilizada nas previsões, tornando os preços propostos pela farmácia no **melhor equilíbrio de exploração do lucro possível.**



Variável dependente do problema

Preço final pelo qual é vendido cada medicamento.

Este valor irá flutuar de acordo com as variáveis independentes relevantes para cada venda.

O mesmo medicamento poderá possuir valores diferentes em cada farmácia visto que cada farmácia **estipula o preço** conforme as suas **vendas anteriores** deste medicamento.

Todo o projeto foi desenvolvido de forma a potenciar o incremento desta variável **sem**, com isso, **afetar a venda do medicamento**.



Variáveis independente do problema

Distância: Um dos principais fatores que faz variar o que um cliente está disposto a pagar ou não por um determinado medicamento é a sua **distância às farmácias que possuem este medicamento em stock.**

Urgência: Do paciente em obter o medicamento, sendo que isto se reflete no facto de o paciente aceitar esperar que uma farmácia **reponha o seu stock** ao invés de **pagar mais** pelo medicamento noutra farmácia.

Época: Implica a altura do ano em que o cliente compra o produto e está condicionada à altura de **mais procura do medicamento.**



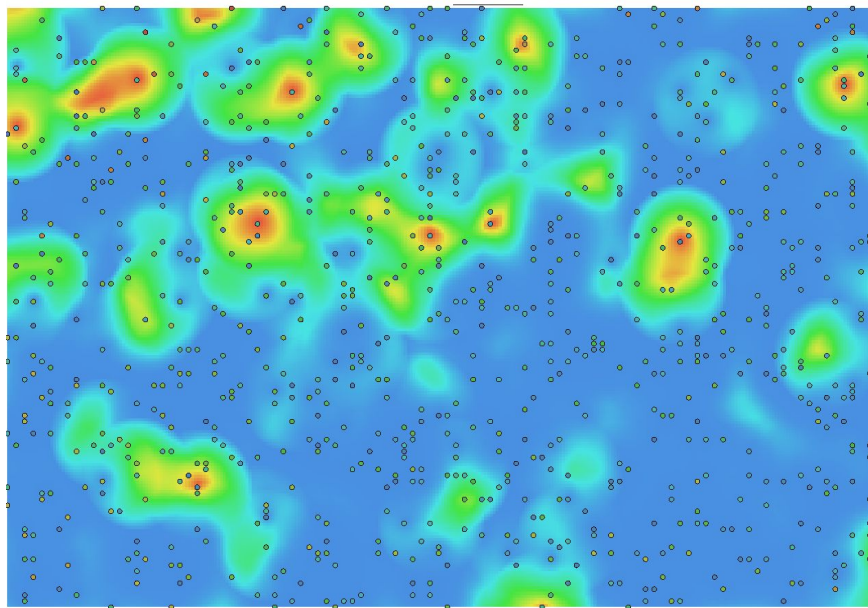
Experiências realizadas

Realizámos algumas experiências com o âmbito de criar uma **base de dados** com informação acerca das **compras efetuadas**. E, através da mesma, poder criar um **modelo de decisão** que permita avaliar e **prever** que preços poderão ser propostos pela farmácia.

Foram ainda gerados modelos classificativos como o Naive Bayes, k-nn para analisar os dados do dataset.



Estatísticas retiradas a partir dos dados obtidos



X-Axis : Client X

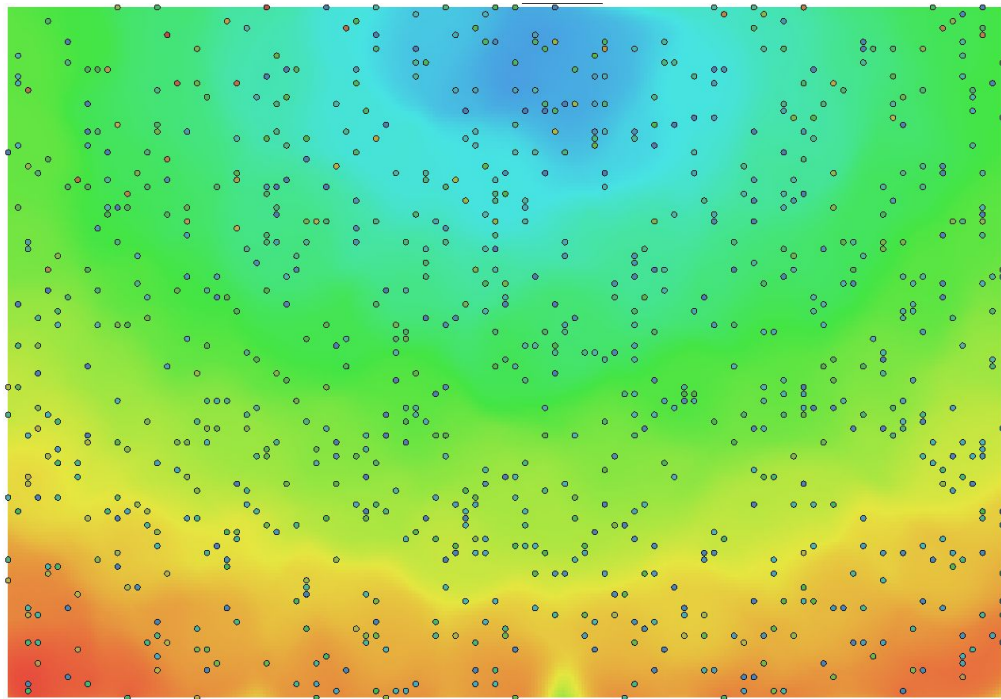
Y-Axis : Client Y

Point Color: OfferPrice

Density: Label



Estadísticas retiradas a partir dos dados obtidos



X-Axis : Client X

Y-Axis : Client Y

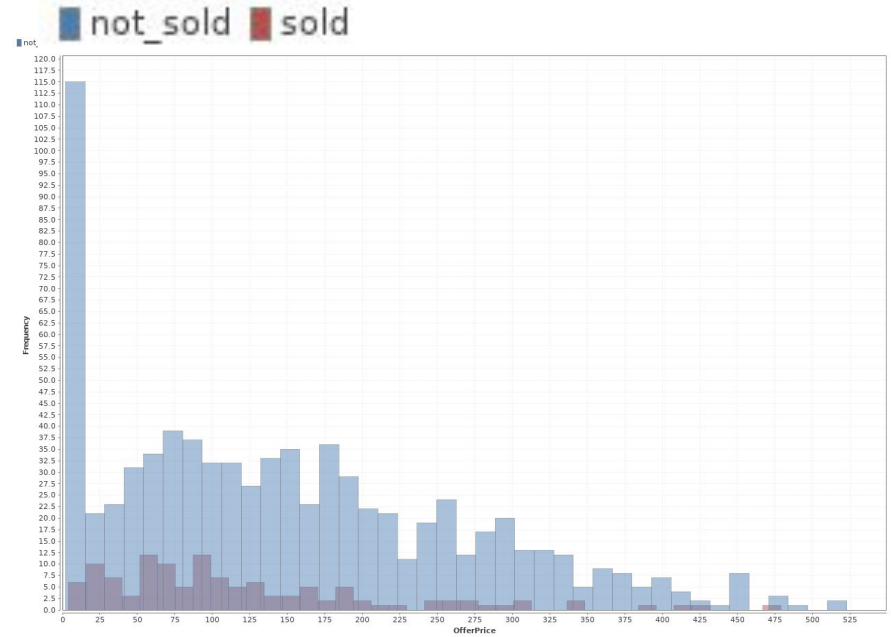
PointColor: OfferPrice Density: distance

Relação entre o preço oferecido e a distancia dos clientes à farmácia

X-Axis : Offer Price

Y-Axis : Label

Como podemos ver e logicamente induzir se o offerprice for muito elevado o cliente naturalmente não vai comprar, mas evidenciou-se um outlier de que se o preço oferecido for muito baixo, o cliente também não compra porque possivelmente noutras farmácias o preço ainda está mais baixo. isto possivelmente é um problema nos agentes que gerou os dados.

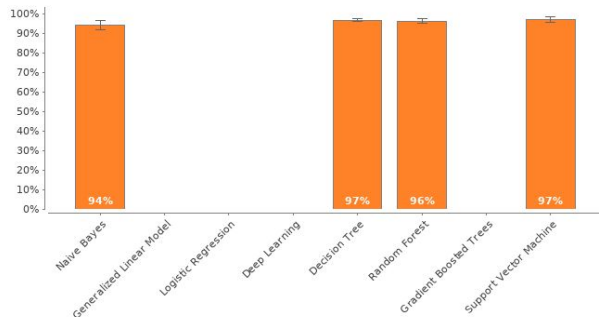




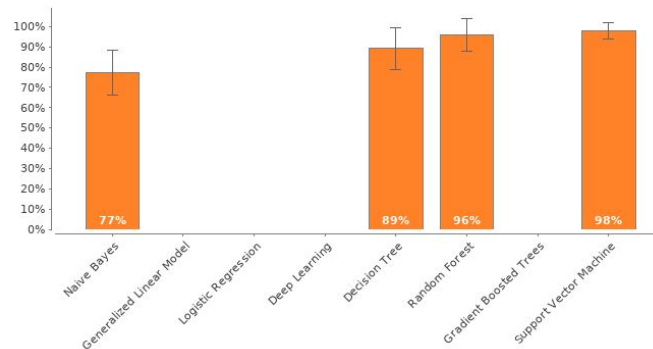
Análise de dados através do RapidMiner

1º começamos por explorar as ferramentas do RapidMiner e começamos com o Auto Model, verificamos que não tínhamos falhas nos dados, mas que tínhamos dados redundantes ou estáticos. Fizemos uma limpeza dos dados e executamos o Auto Model novamente..

Accuracy



Precision



Desenvolvemos alguns modelos com Split Validation para garantir que os modelos eram treinados segundos as normas.

Criamos processo baseados em Decision Tree, Naive Bayes , Knn e Linear Regression, sendo que obtemos os seguintes resultados.

Linear Regression

accuracy: 95.22%

	true not_sold	true sold	class precision
pred. not_sold	228	5	97.85%
pred. sold	8	31	79.49%
class recall	96.61%	86.11%	

Decision Tree

accuracy: 95.96%

	true not_sold	true sold	class precision
pred. not_sold	235	10	95.92%
pred. sold	1	26	96.30%
class recall	99.58%	72.22%	

Naive Bayes

accuracy: 93.01%

	true not_sold	true sold	class precision
pred. not_sold	221	4	98.22%
pred. sold	15	32	68.09%
class recall	93.64%	88.89%	

K-nn

accuracy: 92.28%

	true not_sold	true sold	class precision
pred. not_sold	228	13	94.61%
pred. sold	8	23	74.19%
class recall	96.61%	63.89%	



Análise de dados através do RapidMiner

Com um accuracy muito semelhantes, o modelo de regressão linear e o modelo de Naive Bayes, são bastante eficientes com os dados usados. Os resultados do modelo de regressão linear não surpreendem, uma vez que a fórmula que estamos a usar para os agentes gerarem o preço que vão propor ao Cliente é bastante simples e linear.

Mas com a introdução da distância como peso, a linear regression desceu bastante a fidelidade do modelo, tendo assumindo os seguintes valores:

accuracy: 96.31%

	true not_sold	true sold	class precision
pred. not_sold	25949	890	96.68%
pred. sold	146	1086	88.15%
class recall	99.44%	54.96%	



Problemas

No desenvolver do projecto deparamos-nos com problemas relacionados com a geração de dados nos agentes que não permitiu uma exploração profunda da ferramenta para a criação de modelos mais fiáveis e complexos para otimização do preço manualmente sem usar a ferramenta auto model para ter um finetune e uma melhor experiência de aprendizagem.

A geração dos dados está a ser feita aleatoriamente por cada agente, por isso estes modelos só são válidos no mundo em que foram gerados e não é possível otimizar o preço de um produto e depois testar o modelo em casos reais.



Conclusões

O produto final é capaz de simular compras e vendas entre várias farmácias e vários clientes, onde a farmácia com base em vendas já realizadas, após o pedido do produto por parte do cliente, lhe **faz uma proposta de venda com um preço que pretende maximizar o seu lucro.**

Foi necessário **treinar o modelo**, sendo que, numa fase inicial quanto mais treinado melhor responde de acordo com as circunstâncias, isto é, com uma maior taxa de sucesso nas decisões.

Percebemos, **o poder** de uma base de dados bem povoada, e o que é possível a partir da mesma explorar, representando uma **possível vantagem no mercado** se bem aproveitada para a maioria dos produtos e serviços que existem nos dias de hoje.



Melhoramentos

Poderemos num futuro próximo, testar o modelo com **novas entradas de dados**, de forma a tentar abordar uma **variedade** maior de casos.

Outra alternativa seria abordar também a **competição das farmácias** e não só as transações dos clientes para determinar se existiu ou vai existir uma rotura de stock.

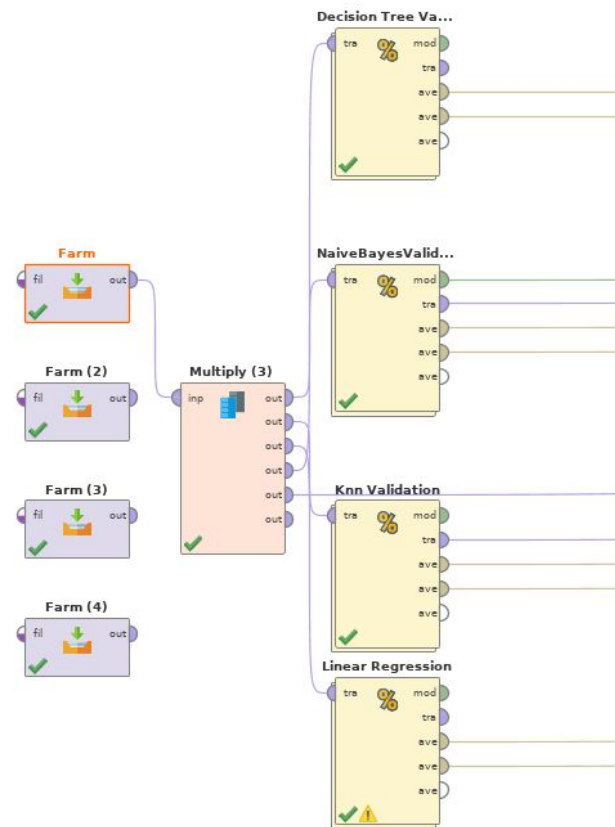
Desenvolver modelos mais complexos.



Anexos - Processos

Cada csv farm tem configurações diferentes no setup das labels e pesos para conseguirmos explorar melhor as diferenças no RapidMiner.

Cada operador é um SplitValidation com o processo relacionado com o modelo no seu subprocesso para conseguirmos validar os modelos com dados de teste.





Informação adicional

O processo principal utilizado para replicar os resultados foi o multitest.

Os restantes processos no ficheiro são os que desenvolvemos no início e as várias experiências e iterações até atingirmos ao resultado final.