



Escola de Engenharia  
Universidade do Minho

Mestrado Integrado em Engenharia de  
Gestão e Sistemas de Informação  
2023/2024

4º Ano

1º Semestre

Aprendizagem Automática em Sistemas de Informação



Francisco Miguel Pinheiro Cardoso

A79570

## Índice

1. Introdução .....	1
2. Data Mining Model.....	2
3. BI System .....	4
3.1 System Architecture .....	4
3.2 Multidimensional Model .....	5
4. ETL .....	5
5. Indicadores .....	6
5.1 Business Objectives .....	6
5.2 Critical Success Factors.....	6
5.3 Measurements .....	7
5.4 KPI'S .....	7
6. Aggregated Data.....	8
6.1 Calculated Fields.....	8
7. DashBoard .....	9
8. Conclusão .....	11

## 1. Introdução

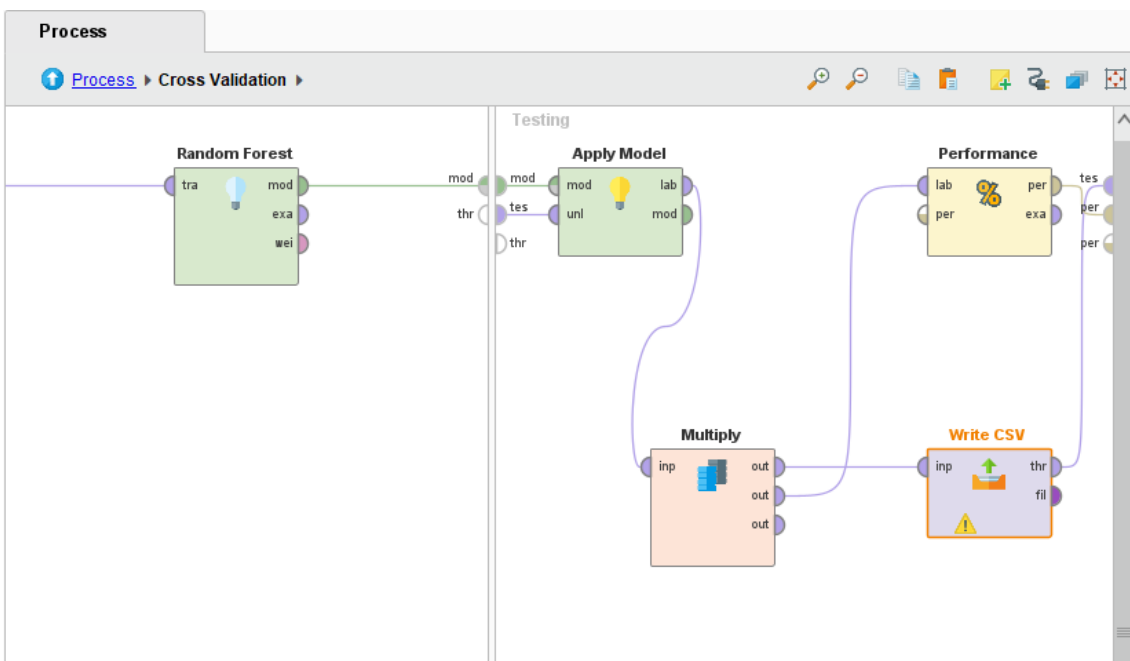
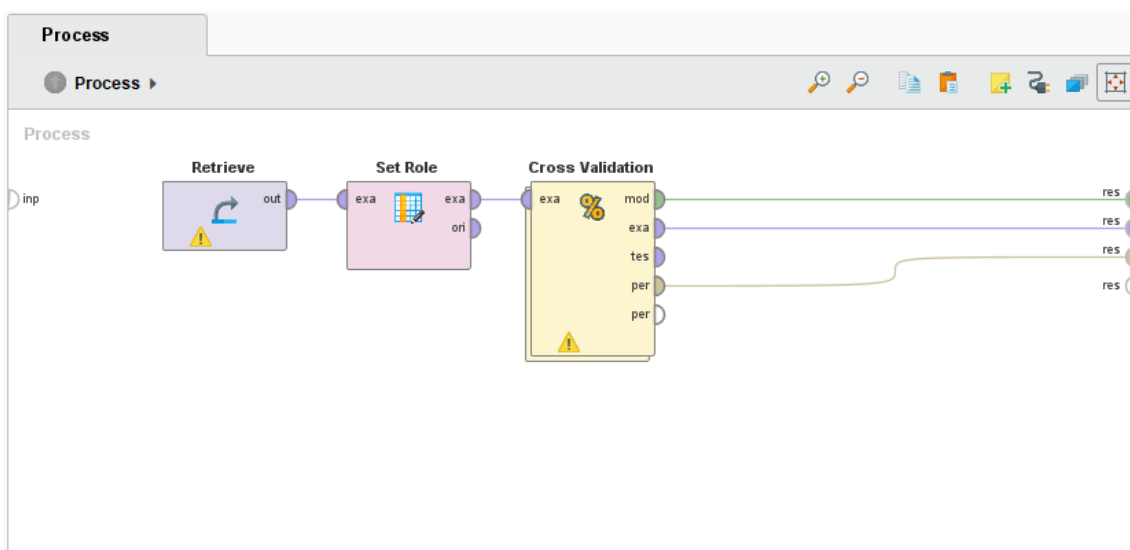
O presente relatório realiza-se no âmbito da unidade curricular “Aprendizagem Automática em Sistemas Empresariais”, lecionada no 1º ano do Mestrado em Engenharia e Gestão de Sistemas de Informação, tendo como intuito a aplicação da metodologia CRISP-DM “Cross Industry Standard Process for Data Mining” para compreensão dos conceitos, princípios e recursos associadas ao Data Mining, onde realizarei a última fase de implementação, produzindo o Business Intelligence.

Para este projeto, tenho como objetivos, obter um componente analítico de Data Mining, através do projeto realizado anteriormente; formular um conjunto de KPI's a partir dos objetivos de negócio e construir dashboards que permitem visualizar a informação relativa ao sistema de BI e os KPI's definidos.

## 2. Data Mining Model

Tendo em conta que o objetivo do presente projeto é a criação de uma plataforma de BI com a adição da componente da Data Mining, foi necessário desenvolver uma estrutura de um modelo de data mining para, posteriormente, construir o modelo multidimensional.

Assim, de acordo com o trabalho anterior, destaquei o cenário que apresentou os melhores resultados: cenário 3, para o modelo random forest, sendo que esse modelo foi novamente corrido em RapidMiner, adicionando um novo componente – o Write CSV- tal como está representado nas seguintes figuras. Este componente permite extrair um ficheiro csv dos diferentes cenários.



Os dados obtidos, para além dos dados que estes datasets originais continham, inclui mais um atributo “predicion (price)”, correspondente ao valor que o model prevê, como podemos visualizar na seguinte imagem.

	A	B	C	D	E	F	G	H	I
1	brand	model	model_year	milage	fuel_type	ext_col	int_col	price	prediction(price)
2	Dodge	Charger SE	2017	63250	Gasoline	Red	Black	18650	41993
3	Tesla	Model S 75D	2016	102260	Gasoline	White	Black	31300	35601
4	Mercedes-Benz	E-Class E 350	2014	35000	Gasoline	Red	Beige	29998	36618
5	BMW	M2 Competition	2020	25013	Gasoline	Black	Black	55000	49994
6	Mercedes-Benz	C-Class C 300 4MATIC Sport	2013	98000	Gasoline	White	Beige	15500	18425
7	Land	Rover Range Rover Sport HSE	2007	220000	Gasoline	White	Beige	5750	10705
8	BMW	Z4 3.0si	2008	59257	Gasoline	Gray	Black	18200	26217
9	Porsche	Macan S	2022	20250	Gasoline	White	Beige	68500	113299
10	Audi	R8 Base	2008	49000	Gasoline	Black	Black	63000	32351
11	Hummer	H2 Base	2005	145000	Gasoline	Green	Black	22500	15544
12	Kia	Niro Plug-In Hybrid EX	2023	9400	Plug-In Hybrid	Black	Black	31500	39512
13	Audi	A4 2.0T Premium	2012	125000	Gasoline	Black	Black	10500	12967
14	Lexus	ES 350 Base	2016	39700	Gasoline	White	Beige	25500	42282
15	Maserati	Ghibli S Q4 GranSport	2020	25721	Gasoline	Blu	Nero	52645	49796
16	Ford	F-250 Lariat	2019	80200	E85 Flex Fuel	White	Black	42500	33603
17	Cadillac	ATS 2.5L Luxury	2013	225000	Gasoline	Blue	Black	13600	15141
18	Mitsubishi	Lancer Evolution MR	2006	128000	Gasoline	Silver	Black	37999	11113
19	Subaru	Ascent Touring 7-Passenger	2021	32250	Gasoline	Black	Black	36000	31354
20	Cadillac	DTS Luxury II	2008	94155	Gasoline	White	Beige	10500	19213
21	Audi	A6 55 Premium	2019	10832	Hybrid	Black	Black	42500	75520
22	Jeep	Wrangler Unlimited 4xe Sahara	2022	10500	Plug-In Hybrid	Green	Beige	49999	44424
23	Jaguar	XK8 Base	2000	40000	Gasoline	Blue	Beige	15400	19816
24	Lexus	LC 500 Base	2018	37000	Gasoline	Black	Red	63997	64093
25	Chevrolet	Corvette Stingray w/1LT	2021	10400	Gasoline	Silver	Black	72900	45759
26	Toyota	FJ Cruiser Base	2014	92000	Gasoline	Beige	Black	27500	28399
27	Tesla	Model X Long Range Plus	2020	43534		Black	Black	65990	47529
28	Dodge	Durango SRT	2020	36000	Gasoline	Gray	Red	62000	32392
29	Mercedes-Benz	Maybach S S 600	2016	108500	Gasoline	White	Beige	63500	33266
30	Honda	Accord Hybrid Base	2021	33523	Hybrid	Lunar Silver	Gray	28748	32334
31	Jeep	Grand Cherokee Laredo	2011	98658	Gasoline	White	Gray	14995	14780
32	Mercedes-Benz	AMG E 63 S 4MATIC	2019	17000	Gasoline	White	Brown	95000	80107
33	Mercedes-Benz	C-Class C 300 4MATIC Sport	2015	73500	Gasoline	White	Black	19000	24757
34	Chevrolet	Cruze LT Automatic	2016	122971	Gasoline	Kinetic Blue	Jet Black	12999	25030
35	Ford	Focus SEL	2017	80134	E85 Flex Fuel	White	Charcoal E	13895	27761
36	Mercedes-Benz	SL-Class SL 550	2011	41314	Gasoline	Red	White	30999	34258
37	BMW	X5 PHEV xDrive45e	2022	31850	Plug-In Hybrid	White	Black	60500	66477

### 3. BI System

#### 3.1 System Architecture



Microsoft Excel – Fonte de Dados



Rapid Miner – Data Mining



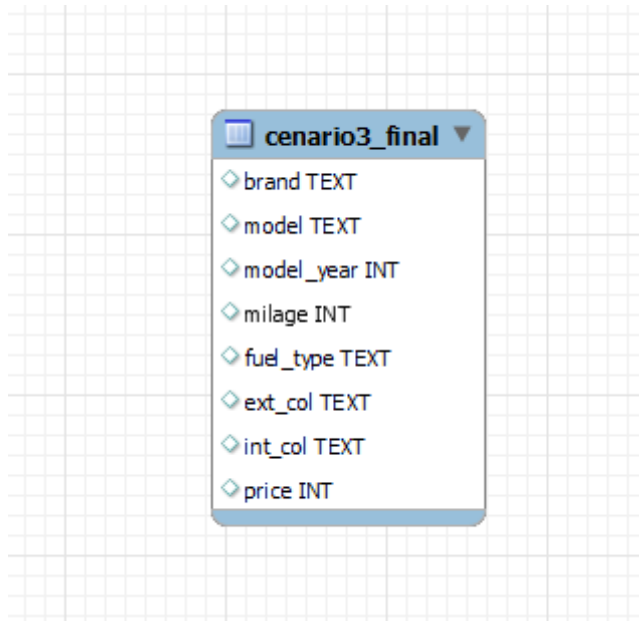
MySQL – Modelo  
Multidimensional



PowerBI – Vizualização

### 3.2 Multidimensional Model

O modelo multidimensional foi desenvolvido com o intuito de lidar com a complexidade do dataset “Cenário 3”, representa uma visão simplificada e estruturada da informação em estudo. Deste modo, não considere que houvesse mudanças a serem necessárias de executar o que nos deixa com o seguinte modelo.



## 4. ETL

Não foi necessário efetuar nenhum tipo de tratamento de dados, uma vez que esse tratamento já tinha sido efetuado no Trabalho anterior, aquando da criação do cenário.

## 5. Indicadores

Os KPI's são medidas que permitem monitorizar o desempenho ao nível dos resultados alcançados em áreas chave das atividades da organização, que são absolutamente críticas para o seu sucesso e expansão.

### 5.1 Business Objectives

De forma a atingir a solução pretendida, é necessário definir objetivos de negócio, para tal, defini os seguintes:

- Apostar na sustentabilidade e promover o negócio com foco nos veículos elétricos
- Aumentar as vendas dos carros mais antigos no website

### 5.2 Critical Success Factors

Os Fatores Críticos de Sucesso (FCS) são as competências ou capacidades centrais que devem ser efetivadas de forma a organização atingir a sua visão.

Objetivo de Negócio	Fatores Críticos de Sucesso
<b>Identificar os veículos elétricos, consoante os valores de preço estimados</b>	Evitar Preço Médio Estimado > Preço Médio
<b>Identificar os veículos mais antigos, consoante os valores de preço estimados</b>	Evitar Preço Médio Estimado > Preço Médio



### 5.3 Measurements

Estas medidas deverão especificar o resultado prático que cada organização e projeto deverá atingir de acordo com os Objetivos de Negócio definidos. Serão igualmente úteis na monitorização e avaliação do processo conducente à concretização desses objetivos.

Fatores Críticos de Sucesso	Medidas
<b>Evitar Preço Médio Estimado &gt; Preço Médio</b>	Preço Estimativa Médio inferior ao Preço Médio

### 5.4 KPI'S

Medidas	KPI
<b>Preço Estimativa médio inferior ao preço médio</b>	Valor do preço médio e valor de preço estimativa médio

## 6. Aggregated Data

### 6.1 Calculated Fields

Através da ferramenta do PowerBI foram criados campos calculados de forma a calcular os KPIs e todos os dados necessários para a sua visualização.

#### Total de Veículos Elétricos

```
1 Nr Eletric Cars = CALCULATE(COUNTROWS(cenario3), cenario3[fuel_type]="Eletric")
```

#### Percentagem de Veículos Elétricos

```
1 % Eletric Cars = [Nr Eletric Cars]/ [Total Nr]
```

#### Preço Médio de Veículos Elétricos

```
1 Average Price Eletric Cars = CALCULATE(AVERAGE(cenario3[price]), cenario3[fuel_type]="Eletric")
```

#### Preço Médio Estimado Veículos Elétricos

```
1 Average Estiamtion Price Eletric Car = CALCULATE(AVERAGE(cenario3[prediction(price)]), cenario3[fuel_type]="Eletric")
```

#### Média de Anos Por Marcas de Carros Elétricos

```
1 Average Year Eletric Cars = CALCULATE(AVERAGE(cenario3[model_year]), cenario3[fuel_type]="Eletric")
```

#### Total de Veículos Antigos (1974 a 2014)

```
1 [Nr Old Cars ] = CALCULATE(COUNTROWS(cenario3), cenario3[model_year]<2014)
```

#### Percentagem de Veículos Antigos

```
1 % Old Cars = [Nr Old Cars ]/ [Total Nr]
```

#### Preço Médio de Veículos Antigos

```
1 Average Price Old Cars = CALCULATE(AVERAGE(cenario3[price]), cenario3[model_year]<2014)
```

#### Preço Médio Estimado Veículos Antigos

```
1 [Average Estiamtion Price Old Car ] = CALCULATE(AVERAGE(cenario3[prediction(price)]), cenario3[model_year]<2014)
```

#### Média de Anos Por Marcas de Carros Antigos

```
1 Average Year Old Cars = CALCULATE(AVERAGE(cenario3[model_year]), cenario3[model_year]<2014)
```

## 7. DashBoard

A Dashboard contém duas páginas que irão apresentar a informação sobre os dois objetivos de negócio alvo.

Na primeira página e na segunda páginas estão disponibilizadas uma opção interativa sobre o total de veículos disponíveis em relação às marcas e modelos, mais por questão adicional informativa, bem como um gráfico que relaciona quantidade de carros com marcas e modelos escolhidos com os anos que têm.

De seguida, de forma a promover a venda de carros elétricos, foi criada uma secção informativa com o nº total de carros elétricos, bem como a percentagem; um valor médio dos carros existentes e o valor médio estimado do preço para novos carros que possam surgir. Apresenta ainda um gráfico informativo sobre a relação total de carros com os anos dos mesmos.



Na segunda página, de forma a promover a venda de carros antigos, foi criada uma secção informativa com o nº total de carros entre 1974 e 2014, bem como a percentagem; um valor médio dos carros existentes e o valor médio estimado do preço para novos carros que possam surgir. Apresenta ainda um gráfico informativo sobre a relação total de carros com os anos dos mesmos.



## 8. Conclusão

Inicialmente, comecei por extrair do RapidMiner o cenário que obtive melhores resultados na fase anterior, com a “prediction” dos preços dos carros.

Posteriormente, optei por não realizar transformações em relação aos dados, por considerar que eram suficientes para os efeitos pretendidos do objetivo de negócio.

Elaborei os KPI's alinhados com os objetivos de negócio definidos anteriormente, no projeto 1. Após definidos os KPI's e tendo como base os mesmos, através do PowerBI, criei os gráficos e tabelas, de forma a apresentar todas as informações que considere relevantes, sobre a forma de dashboard.

O projeto no geral, excedeu as minhas expectativas, melhorei bastante as minhas competências com as diversas ferramentas utilizadas, o que tornou um projeto bastante satisfatório de realizar. Obtive uma experiência prática muito realista de negócio. No entanto, senti algumas dificuldades em relação ao facto de o ter realizado sozinho.

Em suma, sinto que cumpri os objetivos estipulados, apesar de não terem sido satisfatórios a nível de negócio, possivelmente com outras técnicas poderia ter feito melhor.