



Business Intelligence – MEI – University of Coimbra

Análise dos voos nos EUA

<Meta 3>

Equipa "<nome da equipa>":

André Miguel Pereira Pinho, [<apinho@student.dei.uc.pt>](mailto:apinho@student.dei.uc.pt)

Rafael Filipe Pereira Pinho, [<rffpinho@student.dei.uc.pt>](mailto:rffpinho@student.dei.uc.pt)

O projeto

- Este projeto tem como objetivo fornecer dados estatísticos, ratings operacionais e previsões dos voos domésticos nos EUA:
 - Importantes para a gestão global dos transportes pelo DOT
 - Para as companhias aéreas compararem o desempenho do seu serviço com os seus concorrentes

Objetivos

- Prever o número de voos e o número de passageiros para os próximos três anos tanto para as companhias aéreas como para os aeroportos
- Classificar o desempenho operacional das companhias aéreas até ao momento no ano de 2017

Software

- Pentaho data integration (kettle)
 - Inserir novos dados no data warehouse
 - Extrair valores do data warehouse
 - Criar os ficheiros .arff
- Weka
 - Fazer previsões
 - Classificar as companhias aéreas

Fontes de dados

- A informação atual do nosso data warehouse referente aos voos domésticos em 2015, às companhias aéreas e aos aeroportos inseridos na meta anterior
- Site do governo dos EUA que forneceu para esta meta
 - Informação dos voos domésticos referentes aos anos de 2014, 2016 e 2017 (apenas os primeiros três meses) usados para a classificação [1]
 - Também forneceu dados históricos relativos ao número de voos e passageiros para cada companhia aérea e aeroporto nos últimos 10 anos usados para a previsão [2]

[1] https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time

[2] https://www.transtats.bts.gov/Data_Elements.aspx?Data=1

Atributos selecionados para cada análise

Previsão companhias aéreas

- @attribute companhiaAerea string
- @attribute ano numeric
- @attribute nVoos numeric
- @attribute nPassageiros numeric

Previsão aeroportos

- @attribute aeroporto string
- @attribute ano numeric
- @attribute nPartidas numeric
- @attribute nPassageiros numeric

Classificação do desempenho operacional

- @attribute percentagemCancelados numeric
- @attribute percentagemDesviados numeric
- @attribute percentagemVoosAtrasadosP numeric
- @attribute percentagemVoosAtrasadosC numeric
- @attribute tempoMedioAtrasoP numeric
- @attribute tempoMedioAtrasoC numeric

Preparação dos dados

- Executámos o processo ETL desenvolvido na meta anterior para inserir e limpar os novos dados referentes aos anos de 2014, 2016 e 2017
- Usámos o fluxo kettle abaixo para gerar os ficheiros .arff com a informação necessária tanto para as previsões como para as classificações



Algoritmos

- Para a **previsão** usámos o algoritmo 'MultilayerPerceptron' com uma learningRate de 0.01, porque após várias experiências foi aquele que deu os melhores resultados
- Para a **classificação** optámos por usar o algoritmo 'RandomForest', porque após comparar com outros algoritmos foi aquele que deu a melhor percentagem de instâncias corretamente classificadas

Algoritmo de classificação	Use training set (instâncias)	
	Corretamente classificadas	Incorretamente classificadas
OneR	51.05 %	48.95 %
SMO	64.77 %	35.23 %
BayesNet	73.63 %	26.37 %
J48	91.56 %	8.44 %
<u>RandomForest</u>	100.00 %	0.00 %

Algoritmos (continuação)

- Sumário e matriz de confusão para o algoritmo de classificação RandomForest

=== Summary ===

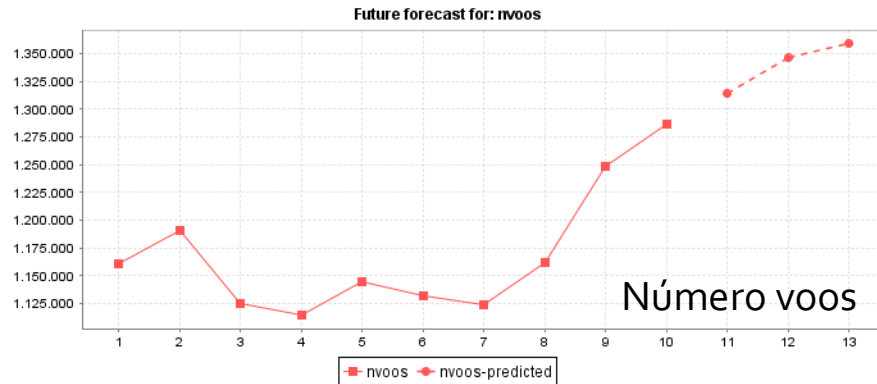
Correctly Classified Instances	474	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0307		
Root mean squared error	0.0731		
Relative absolute error	18.4389 %		
Root relative squared error	25.3709 %		
Total Number of Instances	474		

=== Confusion Matrix ===

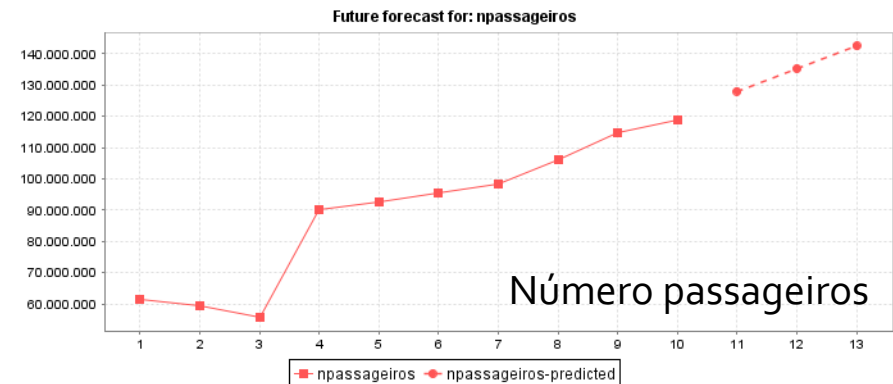
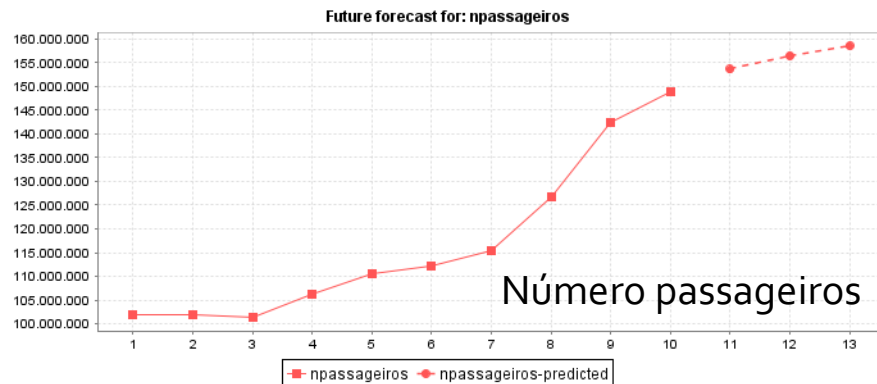
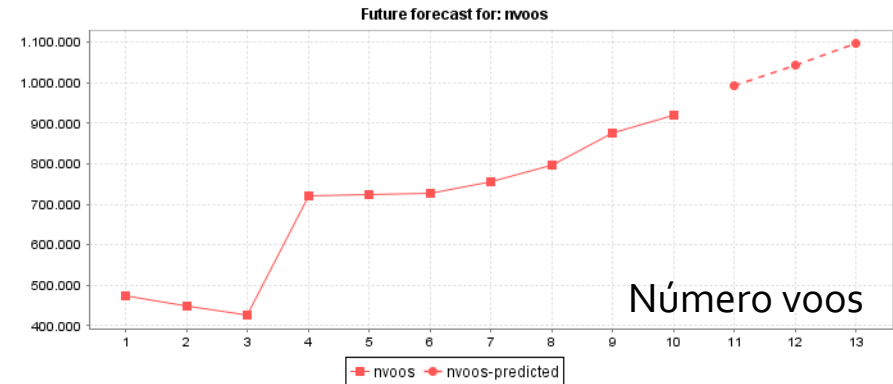
	a	b	c	d	e	f	g	h	i	j	<-- classified as
a = Um	0	0	0	0	0	0	0	0	0	0	
b = Dois	0	8	0	0	0	0	0	0	0	0	
c = Tres	0	0	49	0	0	0	0	0	0	0	
d = Quatro	0	0	0	86	0	0	0	0	0	0	
e = Cinco	0	0	0	0	113	0	0	0	0	0	
f = Seis	0	0	0	0	0	95	0	0	0	0	
g = Sete	0	0	0	0	0	0	71	0	0	0	
h = Oito	0	0	0	0	0	0	0	36	0	0	
i = Nove	0	0	0	0	0	0	0	0	15	0	
j = Dez	0	0	0	0	0	0	0	0	0	1	

Resultados - Previsão companhias aéreas

Southwest Airlines Co.



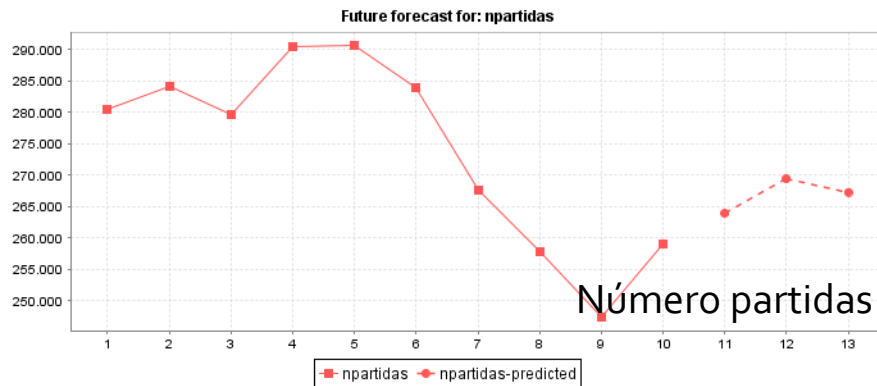
Delta Airlines Inc.



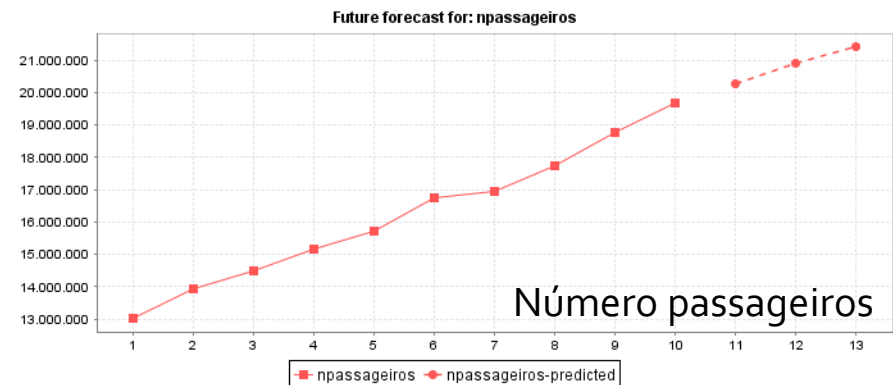
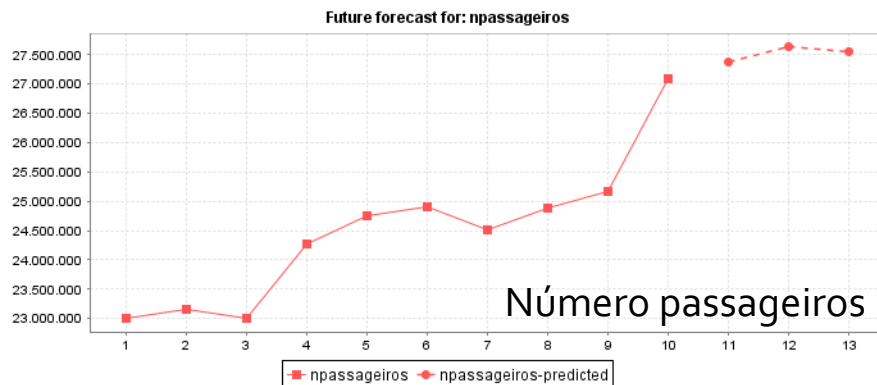
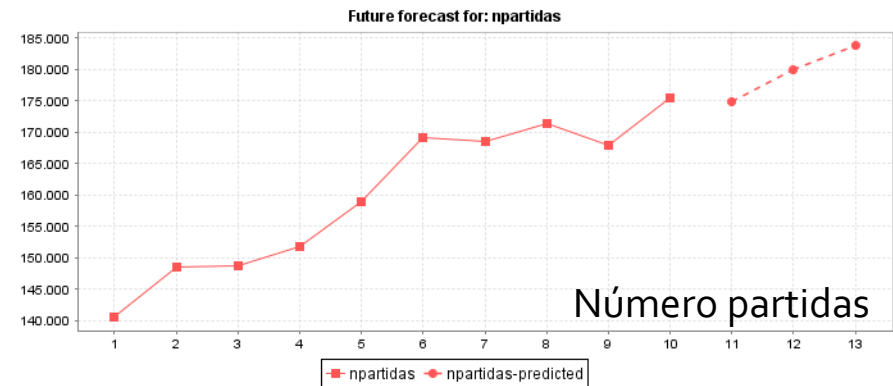
Podemos ver uma tendência crescente no número de voos e passageiros para as duas maiores companhias aéreas de voos domésticos nos EUA, a 'Southwest Airlines Co.' e a 'Delta Airlines Inc.'

Resultados - Previsão aeroportos

Denver International Airport



San Francisco International Airport



No aeroporto de San Francisco o número de voos tem tendência a aumentar nos próximos 3 anos de forma gradual. No aeroporto de Denver a tendência é que aumente em 2017 e 2018 e decresça em 2019

Resultados - Classificação

Nome companhia aérea	Classificação de 2016 (primeiros 3 meses)	Classificação de 2017 (primeiros 3 meses)	
Alaska Airlines Inc.	7	5	↓
American Airlines Inc.	6	7	↑
Atlantic Southeast Airlines	6	5	↓
Delta Airlines Inc.	7	7	—
Frontier Airlines Inc.	7	6	↓
Hawaiian Airlines Inc.	9	8	↓
JetBlue Airways	5	4	↓
Skywest Airlines Inc.	5	4	↓
Southwest Airlines Co.	7	6	↓
Spirit Airlines	5	6	↑
United Airlines Inc.	6	6	—
Virgin America	5	3	↓

A qualidade do desempenho operacional, regra geral diminuiu ligeiramente de 2016 para 2017, em igual período

Questões

