



Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2024/2025

Guião prático nº 14

Março, 2025

Tema

Aplicação de técnicas de aprendizagem com KNIME: Redes Neuronais Artificiais.

Objetivos de aprendizagem

Com a realização desta ficha prática pretende-se que os estudantes:

- Apliquem nodos de aprendizagem baseados em redes neuronais;
- Usem nodos de avaliação de modelos;

Enunciado

Os dados incluídos no *dataset* «*flights_delay.csv*» são relativos a voos realizados nos EUA e respetivas características.

O *target* é a coluna “*delayed*” que indica o atraso no voo.

Realizar as tarefas seguintes:

1. Carregar o *dataset* e aplicar nodos de exploração, preparação e tratamento de dados;
 - a. Há colunas com valores em falta?
E com valores constantes?
E com valores com elevada variância?
 - b. Para cada um dos casos anteriores, aplique os nodos necessários para tratamento de dados;
 - c. Aplicar nodos para visualização de *outliers* e correspondente tratamento.
2. Particionar o *dataset* com estratificação (*stratified sampling*), usando 70% para treino e 30% para teste;
 - a. Explique em que consiste a opção *stratified sampling*;
3. Normalizar o *dataset* de treino usando a normalização Min-Max entre 0 e 1;
4. Normalizar os dados de teste com o modelo de normalização anterior;
 - a. Qual a razão para aplicar o normalizador da alínea 3?
 - b. Poder-se-ia aplicar um nodo de normalização novo? Quais as consequências?
5. Aplicar redes neuronais para treinar o modelo;
6. Aplicar o nodo SCORER para obtenção de métricas de qualidade;



Descrição do *dataset* FLIGHTS_DELAY

ATRIBUTO	DESCRIÇÃO
Year	ano em que o voo ocorreu
Month	mês em que o voo ocorreu
DayOfMonth	dia do mês em que o voo ocorreu
DayOfWeek	dia da semana em que o voo ocorreu
DepTime	hora de partida real
CRSDepTime	hora de partida programada
ArrTime	hora de chegada real
CRSArrTime	hora de chegada programada
UniqueCarrier	ID da transportadora aérea
FlightNum	Número do voo
TailNum	Número da cauda do avião
ActualElapsedTime	tempo real de duração do voo
CRSElapsedTime	tempo programado de duração do voo
AirTime	tempo total no ar
ArrDelay	atraso na chegada
DepDelay	atraso na partida
Origin	aeroporto de origem
Dest	aeroporto de destino
Distance	distância do voo
TaxiIn	tempo no solo após chegada (para estacionar)
TaxiOut	tempo no solo na partida (para descolar)
Cancelled	indica se o voo foi ou não cancelado
CancellationCode	código de cancelamento (razão para o voo ter sido cancelado)
Diverted	indica se o voo foi desviado para outro aeroporto
CarrierDelay	atraso devido a problemas com a transportadora
WeatherDelay	atraso devido a condições meteorológicas
NASDelay	atraso devido a problemas no sistema nacional de aviação
SecurityDelay	atraso devido a questões de segurança
LateAircraftDelay	atraso devido a chegada tardia de outro avião
Delayed	indica se o voo teve ou não atraso (<i>target</i>)

Mais detalhes sobre estes dados podem ser encontrados neste *link*: [kaggle.com/datasets/giovamata/airlinedelaycauses](https://www.kaggle.com/datasets/giovamata/airlinedelaycauses)