

MODELOS de PREVISÃO

TRABALHO INDIVIDUAL de 2ª ÉPOCA

Anabela Costa (anabela.costa@iscte-iul.pt)

1. Enunciado

Desenvolver um modelo que permita estimar o *score* médio de *review* de hóspedes para hotéis no Dubai, na plataforma digital *TripAdvisor*. O *score* no TripAdvisor possui uma escala de 1 a 5.

As variáveis consideradas no *dataset Dubai_data* (**disponível no Moodle, no tópico “Trabalho Individual_2ª Época”**) são as seguintes:

- *score_review* – representa o *score* médio de review de hóspedes para o hotel;
- **CarimboTripAdvisor** – indica se o hotel possui *TripAdvisor Excellence stamp*;
- **Patrocinado** – indica se o hotel é patrocinado;
- **Breakfast** – indica se o hotel tem pequeno-almoço incluído;
- **WiFi_gratuito** – indica se o hotel tem em destaque o *amenitie Wi-Fi* gratuito;
- **Estacionamento_gratuito** – indica se o hotel tem em destaque o *amenitie* estacionamento gratuito;
- **Piscina** – indica se o hotel tem em destaque o *amenitie* piscina;
- **Restaurante** – indica se o hotel tem em destaque o *amenitie* restaurante;
- **Servico_quartos** – indica se o hotel tem em destaque o *amenitie* serviço de quartos;
- **Praia** – indica se o hotel tem em destaque o *amenitie* praia;
- **Bar_lounge** – indica se o hotel tem em destaque o *amenitie* bar/lounge;
- **Tomar_medidas_seguranca** – indica se o hotel anuncia que possui medidas de segurança para com a pandemia da COVID-19;
- **Visitar_website_hotel** – indica se o hotel possui o *link* para o próprio *Website*;
- **logavaliacoes** – representa o número de *reviews* já escritas, por hóspedes, para o hotel (valores normalizados pelo logaritmo natural).

2. Recomendações

O desenvolvimento do projeto e do relatório final deve seguir a metodologia CRISP-DM, sendo recomendadas as seguintes etapas:

Etapas 1 – Com base nos dados e na pesquisa realizada sobre os dados (*Business Understanding*), definir o(s) problema(s)/questões que pretendem resolver/responder e justificar a escolha das variáveis.

Etapas 2 – Fazer a limpeza dos dados (tratamento de valores omissos, *outliers*, etc).

Etapas 3 – Analisar gráficos e estatísticas descritivas das variáveis.

Etapas 4 – Analisar correlação entre as variáveis.

Etapa 5 – Pré-processamento dos dados e manipulação das variáveis (agrupar, juntar, eliminar, transformar, codificar as variáveis).

Etapa 6 – Dividir o *dataset* em conjuntos de treino e de teste.

Etapa 7 – Usar todos os algoritmos de aprendizagem supervisionada, estudados na Unidade Curricular, que sejam adequados à previsão do preço do quarto.

Etapa 8 – Validar o modelo escolhido e fazer a previsão da variável dependente/alvo (sobre o conjunto de teste).

Etapa 9 – Avaliar a performance da previsão feita (sobre o conjunto de teste).

Etapa10 – Interpretação/ explicação dos resultados obtidos a partir dos dados (prós e contras).

Etapa 11 – O relatório final deve ser submetido no Moodle em formato pdf. Devem ainda submeter um ficheiro zipado com o *dataset* final e o ficheiro com o código desenvolvido. **A data limite para a submissão é 28 de janeiro (até às 23h59m).**

Etapa12 – O código deve estar devidamente comentado.

Etapa13 – Deverá incluir todas as experiências feitas, incluindo a análise e preparação de dados, modelação, e avaliação do modelo.

Etapa 14 – A interpretação dos resultados deve ser efetuada de forma crítica e não apenas factual.

Etapa15 – O relatório tem um limite de 25 páginas.