

1. ENQUADRAMENTO

A empresa OLX é um retalhista *online* português que pretende desenvolver uma ferramenta que permita de uma forma automática classificar a gama de preços dos *smartphones* (ex. baixa, alta). Por sua vez, o *output* desta ferramenta servirá como sugestão da gama de preços mais indicada aos utilizadores que pretendem vender na plataforma os seus *smartphones*. O valor acrescentado para o utilizador, será o aumento de rapidez com que estes conseguem vender os seus artigos. Já para a empresa será a redução do número de artigos que estão listados no seu *website*, o que permite reduzir a carga nos seus servidores e, por consequência, poupar nos custos de infraestrutura.

Para atingir este objetivo, a OLX quer descobrir através dos dados guardados internamente na sua base de dados qual a relação entre as características de um telemóvel (ex. RAM, Memória Interna, etc.) e o seu preço de venda. Adicionalmente, a solução desenvolvida deverá adaptar-se em conformidade com os avanços tecnológicos, atualizando assim a forma como é atribuída a gama de preços para o mesmo modelo de *smartphone*.

Para avaliar os requisitos do projeto, a OLX desafiou os alunos da Pós-graduação de *Business Analytics* a desenvolverem um protótipo que permita simular a nova funcionalidade. O protótipo será crucial para a OLX, pois permitirá aos engenheiros de *software* implementar a nova funcionalidade no *website* da OLX. Como entregáveis, a OLX pretende o seguinte:

- Uma apresentação para demonstrar a solução proposta e todas as fases do projeto. Os alunos deverão ser capazes de explicar a tecnologia utilizada, como também realçar os principais benefícios da solução proposta.
- Um protótipo da ferramenta no *Streamlit* que servirá para simular a nova funcionalidade da OLX.

2. DESCRIÇÃO DOS DADOS FORNECIDOS

0. `smart_id` – Identificador único do *smartphone*.

1. `battery_power` – A capacidade da bateria do *smartphone* é a quantidade de electricidade que uma bateria totalmente carregada pode fornecer a um dispositivo autónomo antes de ser completamente descarregada. Em termos simples, este indicador pode dar uma ideia aproximada de quanto tempo o *smartphone* funcionará por si só, antes de ser completamente descarregado.

2. `blue` – Tem *Bluetooth* ou não. *Bluetooth* é uma norma tecnológica sem fios de curto alcance que é utilizada para trocar dados entre dispositivos fixos e móveis em curtas distâncias, utilizando ondas de rádio UHF nas bandas ISM de 2,402 a 2,48 GHz e construir redes de área pessoal (PANs). É principalmente utilizada como alternativa às ligações com fios para partilhar ficheiros entre dispositivos portáteis próximos e para ligar *smartphone* e música.

3. `clock_speed` - Velocidade a que o microprocessador executa as instruções. Velocidade do relógio é o número de operações que o processador executa por segundo. Quanto maior for, maior é o desempenho do processador. O número de núcleos do processador e o tamanho da *cache* também são importantes. Agora, mesmo os processadores dual-core mais baratos vêm com uma frequência de 3,5 GHz – este é o nível de um computador multimédia ou de jogos da classe média. Se este indicador for maior, a possibilidade de *overclocking* do processador e o número de núcleos também aumenta.

4. `dual_sim` – Tem *Dual Sim* ou não. O termo *Dual Sim* num *smartphone* significa suporte para dois cartões SIM, um dos quais pode usar, por exemplo, para chamadas pessoais e o segundo para trabalho. Muitos *smartphones* modernos suportam dois cartões SIM.

5. `fc` – Câmara frontal em mega *pixels*. A câmara frontal parece um pequeno olho que se encontra na frente do telefone, no mesmo local onde os sensores estão instalados (ou seja, na parte superior). Outros fabricantes não prestaram a devida atenção às características da câmara frontal no *smartphone*, uma vez que dificilmente interessam às pessoas. A câmara frontal foi utilizada exclusivamente para fazer videochamadas.

6. `four_g` – Tem 4G ou não. 4G é uma geração de comunicações móveis com exigências acrescidas. É costume referir-se à quarta geração como tecnologias promissoras que permitem

a transmissão de dados a uma velocidade até 100 Mbps para móveis (com alta mobilidade) e até 1 Gbps para assinantes fixos (com baixa mobilidade).

7. int_memory – Tamanho do armazenamento físico interno em *Gigabytes*. O armazenamento interno de dados num *smartphone* é onde se encontram dados importantes: o sistema operativo (SO), aplicações instaladas, fotos, vídeos, documentos e outros ficheiros.

8. m_dep – Profundidade do *smartphone* em cm.

9. mobile_wt – Peso do *smartphone* em gramas.

10. n_cores – Número de núcleos do processador. O número total de núcleos num único processador de um *smartphone Android* é tipicamente oito (a maioria das atualizações do *iPhone* tem seis).

11. pc – Resolução da câmara principal em *megapixels*. O número de *megapixels* de um sensor de câmara descreve a resolução de imagem que pode ser capturada. Por exemplo, as câmaras com um sensor de 12 *megapixels* podem tirar fotografias com uma resolução de 4200x2800 *pixels*, uma câmara de 8 *megapixels* permite tirar fotografias com uma resolução de 3264x2468 *pixels*.

12. px_height – Resolução *pixels* em altura.

13. px_width – Resolução *pixels* em largura.

14. ram – Memória de Acesso Aleatório em *Megabytes*. *Random Access Memory* (RAM) é a ligação entre o processador e o sistema de reprodução, porque contém informação temporária necessária para a execução de aplicações.

15. sc_h – Altura da tela do *smartphone* em cm.

16. sc_w – Largura da tela do *smartphone* em cm.

17. talk_time – Duração da bateria com carga completa em chamada (em horas).

18. *three_g* – Tem 3G ou não. A comunicação móvel da terceira geração é construída com base na transmissão de pacotes de dados. As redes da terceira geração 3G funcionam na fronteira do decímetro e estão ligadas à rede. Permitem-lhe organizar uma videochamada, ver filmes e conteúdos individuais no seu telemóvel.

19. *touch_screen* – Tem ecrã tátil ou não. Um ecrã tátil, de facto, é um vidro tátil que funciona de acordo com um esquema simples: tocar no observador permite-lhe realizar quaisquer funções ou sintomas de exposição.

20. *wifi* - Tem *wifi* ou não. *Wi-Fi* é uma tecnologia de rede sem fios que permite a dispositivos como computadores portáteis e de secretária, dispositivos móveis (*smartphones* e artigos de vestuário) e outros equipamentos (impressoras e câmaras de vídeo) acederem à Internet.

21. *price_range* – Esta é a variável alvo com valor 0 (baixo custo), 1 (custo médio), 2 (custo elevado) e 3 (custo muito elevado).

3. OBJETIVOS PRINCIPAIS DO PROJETO

- Simular durante as aulas o desenvolvimento de um projeto de *Machine Learning* em contexto empresarial com recurso à *Cloud*;
- Simular aplicação da metodologia de *Machine Learning pipeline* através de uma aplicação prática com dados reais;
- Evidenciar as principais dificuldades encontradas no desenvolvimento de projetos de *Machine Learning*;
- Capacitar os alunos das competências técnicas base que permitam potenciar o sucesso de projetos de *Machine Learning*.

4. ENTREGA

O Projeto Final é elaborado em equipas de 3 elementos. As equipas de trabalho serão acompanhadas pelo Prof. Luís Dias durante o decorrer das aulas. Cada equipa terá de entregar até ao dia 13 de junho de 2025 o código em *Python* que permita correr o protótipo desenvolvido pelo grupo. Além disso, deverá incluir o código para treinar o modelo de *Machine Learning* no *SageMaker*, o código que permita gerar as previsões em tempo real no *SageMaker* e, por fim, o código que permita executar a aplicação *Streamlit*.

As equipas terão também de fazer uma apresentação do seu projeto, via *Zoom*, no dia 13 de junho de 2025. A apresentação terá de ser de 15 minutos que serão seguidos de 10 minutos para questões. Todos os entregáveis do trabalho deverão ser submetidos no *Moodle* dentro do prazo estabelecido. Em casos de incumprimento, cada dia de atraso corresponderá a uma dedução de 2 valores na nota final do trabalho.

A avaliação do Projeto Final inclui a componente da apresentação (50 %) e a componente do protótipo em *Streamlit* (50 %). O trabalho contém ainda um componente bónus opcional que, no caso de ser entregue, acresce o valor da nota final. Esta componente extra requer a entrega de cada caso de aplicação apresentado nas aulas até ao dia 15 de junho de 2025. Por cada caso de aplicação submetido no *Moodle* que cumpra os requisitos definidos irá ser valorizado em 0.5 valores. No total, irão ser apresentados 4 casos de aplicação nas aulas.