INTRODUÇÃO

INTELIGÊNCIA ARTIFICIAL NA ERA DO E-COMMERCE: TRANSFORMANDO DADOS EM INSIGHTS

1 EMPRESA MELHORES COMPRAS LTDA

1.1 Momento Atual

A Melhores Compras percorreu um caminho de transformação, impulsionada pela exploração de dados estruturados e não estruturados, pela adoção de *cloud computing* e pela imersão de soluções envolvendo *Big Data* e *Business Intelligence*.

Essas tecnologias aplicadas de forma adequada permitiram à empresa se tornar uma das líderes nesse segmento de mercado, garantindo operações otimizadas e entregas rápidas, como também fomeceram *insights* valiosos que transformaram decisões estratégicas em resultados concretos, culminando em uma era de prosperidade sem precedentes.

Os executivos da Melhores Compras, ao verem os frutos desse investimento, compreendem que a aposta em dados foi um divisor de águas.

Não apenas a Melhores Compras se transformou como empresa, mas seus funcionários também cresceram ao promover e auxiliar em uma incrível mudança de paradigma, moldando a Melhores Compras em uma empresa que utiliza seus dados da melhor maneira possível, ajudando na criação de valor e na tomada de melhores decisões.

Evidentemente, apesar de a Melhores Compras ter chegado à sua maturidade no uso de dados, os desafios não pararam, de jeito nenhum. Agora, com os fundamentos concluídos, é hora de normalizar a utilização de dados no dia a dia da empresa, preservando o que foi conquistado até aqui, além, é claro, de buscar mais oportunidades para a companhia.

Pensando nisso, o próximo desafio virá bem a calhar...

1.2 Primeiro Desafio: Criando modelos preditivos (MLDS)

O sonho de consumo de qualquer executivo de vendas é relativamente simples: prever o que irá acontecer em um momento futuro com o volume de vendas da companhia. Com uma previsão assertiva, a empresa pode se preparar para momentos de contração na economia, como pode também aproveitar momentos de expansão econômica para alavancar ainda mais a sua força de vendas.

Quase nunca as vendas de uma empresa dependem exclusivamente dos seus esforços internos. É bastante comum que o sucesso comercial da empresa dependa bastante de critérios macroeconômicos, como as ações de governos, bancos centrais e outros atores que atuam sobre a economia de um país.

Como o estudante de Ciência de Dados já deve ter visto empiricamente, um efeito (no caso, o nível de vendas da companhia) raramente tem uma única causa, e captar a maior quantidade de informações sobre o efeito pode ajudar de forma decisiva em uma previsão.

Sendo assim, os executivos da área comercial desejam que o time de Data Science estude o assunto para buscar, da melhor forma possível, qual será o nível de venda da empresa em uma data futura.

A base de dados que o seu time tem em mãos (**sales.csv**) contém as informações, conforme mostra o Quadro 1.

Date	Indica a data em que a observação foi feita
Weekly_Sales	As vendas da empresa na semana, indicada pelo campo **Date**
Holiday_Flag	Indica se houve um feriado na semana medida
Temperature	A temperatura média registrada naquela semana
Fuel_Price	Preço do combustível. Deve ser dividido por 1.000 para se obter o preço em reais e centavos
CPI	Índice que indica o nível de atividade econômica da região
Unemployment	Nível de desemprego medido nacionalmente. Deve ser dividido por 1.000 para se obter a informação em percentual

Quadro 1: Dicionário de dados do dataset sales.csv

Fonte: Elaborado pelo autor (2025)

De posse dessas informações, a missão do seu time é buscar o **melhor** algoritmo possível para prever a informação **Weekly_Sales** em um momento futuro.

Além do algoritmo para realizar essa previsão, o seu time também deve defender a escolha do algoritmo utilizado. Ou seja, uma justificativa deve ser formulada para o uso do algoritmo indicado. Ao defender a escolha do seu algoritmo, analise quais informações foram mais e menos relevantes para o seu modelo. Que outras informações seriam interessantes acrescentar a uma nova versão do seu trabalho para melhorar a assertividade? Os resultados encontrados são satisfatórios? Por quê?

Lembre-se de que o seu time já tem bastante experiência, logo, o repertório de técnicas e algoritmos dominados é bastante grande. Utilize todo o conhecimento conquistado até aqui para encontrar o melhor modelo preditivo e defender a sua decisão técnica.

```
In [20]: from sklearn.linear_model import LinearRegression
         from sklearn.preprocessing import PolynomialFeatures

         model_name = 'Polynomial Linear Regression'

         polynomial_features= PolynomialFeatures(degree=3)
         pLRegressor = LinearRegression()

         pLr_model = Pipeline(steps=[('polyFeature',polynomial_features ),('regressor', pLRegressor())]

         pLr_model.fit(X_train,y_train)

         y_pred_pLr= pLr_model.predict(X_test)

         print(pLRegressor)

         LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

Figura 1: Exemplo de código de uma regressão linear
Fonte: developer.ibm.com

A entrega pode ser tão simples quanto a Regressão Linear apresentada na Figura 1. O desafio maior é o grupo conseguir **argumentar tecnicamente**, explicando por que essa poderia ser o melhor modelo para resolver o problema. Boas argumentações e explicações técnicas, aliadas a um bom código, garantem uma boa nota nessa atividade. Um bom código, mas com uma argumentação fraca, vai causar uma nota não tão boa quanto o esperado.

Reforçando, para não deixar dúvidas: no mercado de trabalho, isso é o esperado do Cientista de Dados. Capacidade de entender e resolver um problema, além de segurança técnica para explicar o que, como e por que o problema foi resolvido de uma determinada forma.

1.3 Segundo Desafio: Apresentando modelos preditivos (MLDS)

Após ter determinado (espero que com sucesso!) o melhor algoritmo para realizar a previsão desejada; a missão do seu time é comunicar isso de forma eficiente para o *board*.

Lembre-se: nem todos os executivos têm o conhecimento amplo de dados que o seu time tem. Assim, dizer que um modelo tem uma acurácia de 0,26 pode ser (bastante) ineficaz se a sua audiência não sabe o que significa acurácia e o que representa o número 0,26.

Utilize textos, gráficos, fórmulas, *dashboards*, apresentações, enfim, qualquer ferramenta que ajude o seu time a transmitir ao *board* da empresa os resultados do seu trabalho. O *board* da empresa certamente não conhece termos nem definições técnicas do seu dia a dia. Sendo assim, os conceitos técnicos devem ser explicados antes de utilizados.

Veja também que o conceito de **melhor algoritmo** é bastante difuso para um leigo. Sendo assim, o seu time deve deixar bastante claro, em termos didáticos, por que o algoritmo encontrado é o melhor.

Uma excelente ideia mal comunicada pode ser desastrosa, então, bastante atenção e cuidado com essa parte do desafio.

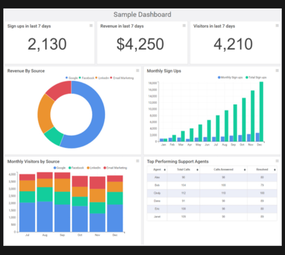

Figura 2: Imagem exemplo de como um dashboard pode ser apresentado.
Fonte: ubiq.co

Em síntese, para garantir uma boa entrega, é essencial que um leigo consiga ler e entender a apresentação dos seus resultados. **Reforçando**: o PBL não exige que seja feita uma *dashboard* tal como ilustrada na imagem acima (embora não seja má ideia). O que o PBL pede é para que você aplique tudo o que aprendeu até agora, dê uma boa polida na sua capacidade de comunicação e encontre a melhor forma de comunicar o resultado do seu trabalho. A criatividade (que às vezes é um atributo que falta para o profissional de Tecnologia da Informação) tem que ser explorada nessa entrega.

1.4 Resumo

De forma breve, os desafios são resumidos no Quadro 2.

Desafio	Resumo das Tarefas
Primeiro Desafio – Criando modelos preditivos (MLDS)	Desenvolver um modelo preditivo para estimar as vendas semanais da empresa *Melhores Compras* usando a base de dados fornecida (contendo variáveis como feriados, temperatura, preço do combustível, CPI e desemprego). É necessário escolher e justificar o algoritmo utilizado, avaliando a relevância das variáveis e discutindo melhorias possíveis para a assertividade do modelo.
Segundo Desafio – Apresentando modelos preditivos (MLDS)	Após definir o melhor algoritmo, deve-se comunicar os resultados ao *board* da empresa de forma clara e acessível. É preciso transformar conceitos técnicos em explicações didáticas, utilizando textos, gráficos, dashboards ou outras ferramentas de visualização, para garantir que até leigos compreendam o impacto e a importância do modelo.

Quadro 2: Resumo dos dois desafios do PBL

Fonte: Elaborado pelo autor (2025)

2. RECOMENDAÇÕES

Essa é a última atividade do curso e, para representar bem o desafio do mercado de trabalho, o escopo da atividade está bastante aberto. Isso significa que há pouquíssimo direcionamento sobre que caminho percorrer para entregar o melhor trabalho possível.

A analogia com as suas futuras (ou atuais?) atividades profissionais é bastante evidente: não é raro que a companhia apresente problemas e o profissional de TI pesquise, estude e traga soluções inovadoras sem que ele tenha sido previamente direcionado.

Sendo assim, é fundamental que você se recorde de tudo o que foi estudado nos últimos meses para lhe ajudar a fazer a melhor entrega possível.

Mas isso não significa que não possamos dar algumas dicas, não é verdade?

Alguns pontos que você deve prestar atenção:

1. **Realize a análise exploratória dos seus dados com bastante cuidado e antes de pensar no algoritmo que deve ser desenvolvido. Existem algoritmos com melhor ou pior performance, a depender da qualidade dos dados. Além disso, a base de dados pode ter defeitos. Se houver defeitos, é importante que seu time os corrija, se possível, e os aponte, se necessário;**
2. **Analise também a variável dependente. Isso deve afunilar bastante as possibilidades de algoritmos que podem ser utilizados nesse problema;**
3. **Quando justificar a escolha de um algoritmo, procure na teoria pontos que reforcem os seus argumentos;**
4. **Se estiver em dúvida sobre mais de um algoritmo, faça testes. Faça a modelagem em diferentes modelos, adote uma métrica e faça comparações. Não se esqueça que muitos algoritmos só conseguem entregar bons resultados após um ajuste fino dos seus hiperparâmetros;**
5. **Na apresentação dos seus resultados, é fundamental o didatismo. No mercado, será bastante comum que você se veja em situações em que deve explicar conceitos técnicos complexos com analogias de simples e fácil entendimento. Um profissional brilhante pode ser facilmente ofuscado se não tiver uma capacidade de comunicação adequada para apresentar os seus achados;**
6. **Ainda falando sobre a apresentação, neste momento do curso, você já tem inúmeras opções para apresentar resultados, com gráficos, *dashboards*, ferramentas de DataViz, entre outros. Busque formas eficientes de transmitir os seus *insights*.**
7. **Ao fazer a sua entrega, observe e leia o material que você produziu. Se essa atividade fosse uma atividade profissional, qual a impressão que você daria para o seu gestor?**

3. ENTREGAS ESPERADAS

a) Arquivo componentes.txt contendo o nome do Grupo, RM e nome dos Integrantes (em ordem alfabética).

b) Arquivo .ipynb contendo o algoritmo que foi utilizado para realizar a previsão de receita futura da companhia.

c) Arquivo .pdf com a explicação do time para o algoritmo escolhido, bem como as suas opções de modelagem. É absolutamente necessário que a explicação seja convincente e o processo de avaliação seguirá o critério de Revisão por Pares, ou seja, **o tutor avaliará vocês como profissionais e não como alunos**.

d) Apresentação das suas conclusões para o *board*. Procure entregar em um arquivo .pdf mas, caso entenda ser necessário outro tipo de mídia para apresentar a sua ideia, sinta-se livre para enviar arquivos em outros formatos.

Gere um arquivo **<nome_grupo>_PBL_TBDO_2o_Ano_Fase7.zip**, agrupe os entregáveis acima e faça o upload conforme orientação da atividade.

Atenção: não serão aceitas entregas por meio de links. A entrega deve ser realizada conforme instruções e anexo, conforme arquivos indicados.

O universo de ciência de dados é imenso e com muitas possibilidades, e o melhor é que estamos melhorando a cada dia. Espero que esteja tudo mais claro agora. Bons estudos e não se esqueça, se precisar de algo, procure o seu tutor para pronto atendimento.

O time do e-commerce das Melhores Compras deseja um bom desafio ao grupo.