



*Big Data Analytics com R e Microsoft Azure Machine Learning 3.0*

# Big Data Analytics com R e Microsoft Azure Machine Learning Versão 3.0

## Lab 3

### Machine Learning em Marketing Digital Prevendo a Probabilidade de Conversão do Lead



**Formação Cientista de Dados 3.0**

**Big Data Analytics com R e Microsoft Azure Machine Learning**

**Lab 3**

**Machine Learning em Marketing Digital  
Prevendo a Probabilidade de Conversão do Lead**

Como Cientista de Dados, você concluiu com sucesso o projeto anterior (Lab 2). Mas outro desafio está por vir.

Seu chefe convida você para uma reunião e informa que a área de Marketing Digital ficou muito satisfeita com seu projeto de previsão do número de usuários convertidos. Mas agora, eles desejam um modelo que entregue uma previsão um pouco diferente.

Considerando detalhes de uma nova campanha de Marketing, eles desejam saber se um lead vai ou não ser convertido (ou seja, se um potencial cliente vai ou não se tornar cliente e comprar o produto). Além disso, eles desejam saber a probabilidade. Eles precisam responder esta pergunta de negócio: Se a previsão do modelo é que um lead será convertido, qual a probabilidade de isso acontecer?

Seu chefe descreve para você que esse é um projeto importante, de grande visibilidade para a área de Analytics e que está um pouco preocupado se você tem as habilidades necessárias para conduzir o projeto e entregar o resultado.

Você diz ao seu chefe que concluiu uma Formação na Data Science Academy (DSA) e se sente apto e confortável com o novo desafio. Seu chefe conhece a qualidade e alto nível dos cursos da DSA (ele mesmo já fez alguns cursos) e diz que assim fica mais tranquilo e que você pode começar o projeto, que deve ser entregue dentro de 3 dias em uma apresentação que vai contar com a presença de todo o departamento de Marketing e alguns membros da diretoria.

Você inicia o projeto, coleta a amostra de dados, aplica seu conhecimento para: interpretar os dados, fazer análise exploratória, pré-processar os dados (que têm variáveis

categóricas), construir algumas versões do modelo, avaliar o modelo usando diferentes métricas e ao final ainda criar um procedimento para o deploy do modelo. Tendo estudado na DSA através de diversos projetos, labs e estudos de casos passo a passo e sempre com foco na solução de problemas de negócio, você não tem qualquer problema em concluir o projeto da área de Marketing Digital.

Entretanto, ao analisar os dados, você se depara com uma questão que o deixa incomodado. Uma das variáveis do dataset apresenta uma informação que tornaria o modelo de Machine Learning enviesado a uma característica que poderia causar discriminação. O dilema está claro e você precisa tomar sua decisão e justificá-la.

Conforme aprendeu na DSA, você não quer emitir uma opinião baseada em achismo, mas sim em dados. Decide então consultar a Lei Geral de Proteção de Dados Pessoais (LGPD) e o Guia de Boas Práticas da LGPD, a fim de identificar as implicações jurídicas do dilema e então tomar sua decisão sobre o que fazer com base nos dados e sob à luz da lei. Você então toma uma decisão e apresenta sua solução para o projeto durante a reunião. Todos parecem satisfeitos com a solução proposta.

Concluída a apresentação você retorna a sua mesa já pensando em como pode melhorar o projeto e como o modelo será periodicamente retreinado. Então o telefone toca! É seu chefe chamando você para outra reunião, desta vez com o board executivo da empresa.

Na reunião você é informado que seu projeto causou impacto, ficou muito bom tecnicamente, mas que ficaram realmente impressionados com a sua postura em identificar o dilema claramente presente nos dados, pesquisar pelas implicações jurídicas e assim evitar que a empresa pudesse ter problemas no futuro. Com base nisso, informam que você será promovido e terá um aumento de 42% efetivo já no próximo mês.

Você retorna a sua mesa e fica feliz e satisfeito não apenas por ter concluído o projeto, mas também por ter tomado a decisão certa.

O cenário descrito acima é exatamente o que você encontrará agora no novo Lab 3 do curso, que traz detalhes sobre a LGPD.

Esse cenário é um caso verídico e foi reportado por e-mail por um dos alunos da DSA.

Boa aula.