

Curso: MBA em Machine Learning in Production (MLP)

Módulo: Introdução ao Machine Learning

Componente: Aprendizado Não-Supervisionado

Aluno: Pedro Luis B. de Almeida

ATIVIDADE 1 – RESPOSTAS

Cenários – Respostas

Cenário 1: Uma empresa de telefonia possui muitos problemas na retenção de clientes. Atualmente, ela não possui nenhuma forma de atuar de forma reativa na retenção de clientes propensos a abandonar seus serviços. Essa rede de telefonia nos procurou no intuito de criar um sistema que indique em quais usuários deveria ser realizada alguma intervenção ou campanha, utilizando uma base histórica de clientes.

R: Aprendizado Supervisionado, pois temos conhecimento, a partir da base histórica, se algum cliente já abandonou o serviço e também as características que podem agregar na procura de um padrão de clientes propensos a desistir do serviço. A melhor técnica aqui seria o uso de algum algoritmo de Classificação, para rotular clientes propensos a desistir ou não. Os dados mais importantes aqui seria o tempo média de vida do cliente, valor pago pelos clientes, tipos de plano ou serviço, NPS dos clientes e etc.

Cenário 2: Uma construtora mantém um conjunto de dados dos imóveis de diversos bairros da cidade de São Carlos, com algumas características do imóvel e o seu valor de venda. O processo de análise de valor do imóvel é um gargalo na operação, pois demanda muito tempo. Portanto, essa construtora quer criar um modelo que ajude a estimar o valor de imóveis ainda não avaliados em função dos seus atributos.

R: Aprendizado Supervisionado, pois a construtora tem uma base de dados com imóveis já precificados e seus atributos. Com isso, podemos estimar com uma Regressão Linear por exemplo o valor do imóvel com base nos dados históricos e atributos de imóveis semelhantes. Os dados importantes aqui seria o tamanho do terreno, tamanho da casa construída, quantidade de quartos e de banheiros, localização e etc.

Cenário 3: Uma rede de lojas de departamento deseja expandir seus negócios em São Carlos. Ela já possui duas lojas e deseja abrir mais três. Para isso, ela gostaria que fosse implementado uma solução que indicasse quais os melhores pontos da cidade para alocar essas novas lojas utilizando, por exemplo, dados demográficos.

R: Aprendizado Não-Supervisionado. Com os dados demográficos das regiões da cidade, conseguimos segmentar os bairros de acordo com o tipo de população que lá habita, criando assim grupos (clusters) que possam orientar qual tipo e tamanho de loja atende melhor cada região. Os dados importantes aqui são o índice demográfico da região, habitantes por quilômetro quadrado, renda per capita, média de idade, índice de óbitos e etc.

Cenário 4: O São Carlos Futebol Clube irá jogar a final da libertadores na próxima semana, porém seu adversário ainda será decidido pelo vencedor do jogo Palmeiras e Corinthians. Para otimizar os treinos da semana, o São Carlos FC gostaria de saber qual time tem a maior probabilidade dele enfrentar, portanto fomos convocados para criar um modelo para prever o resultado deste jogo.

R: Aprendizado Supervisionado. Aqui se trata de um problema de classificação, onde pode ser usado por exemplo, o classificador de NaiveBayes por lidar com probabilidade baseado em variáveis independentes entre si. Os principais dados analisados aqui e que seria ideal contar com dados sobre quantidade de jogos vencidos, empatados e

perdidos, posse de bola, quantidade de gols feitos e sofridos, média de nota dos jogadores e etc.

Cenário 5: Uma universidade está tendo muita dificuldade para redirecionar as reclamações dos alunos para os departamentos corretos. Hoje não existe nenhum tipo de classificação: a reclamação cai para um profissional que lê a mensagem e a redireciona de acordo com seu conhecimento empírico. Para generalizar e agilizar esse processo, a universidade necessita de um sistema que consiga detectar de forma automática para onde aquela mensagem deveria ser redirecionada.

R:Aprendizado Não-Supervisionado. Aqui seria ideal usar a técnica de agrupamento, pois para cada tipo de e-mail, teríamos uma label que poderia o identificar e assim o algoritmo era treinado a partir de assuntos, palavras-chave e os encaixaria em grupos que seriam os departamentos de cada universidade.

Cenário 6: Pense em um cenário, seja de sua empresa ou que você veja como um problema a ser resolvido da sociedade e responda às três perguntas de como ele poderia ser resolvido (parcialmente ou totalmente).

R: Um problema clássico que muitas empresas do ramo financeiro enfrentam são os clientes inadimplentes em relação a empréstimos, créditos, financiamentos e etc. Um método muito utilizado para prever clientes inadimplentes de uma instituição financeira se dá pelo uso do Aprendizado Supervisionado, mais precisamente algoritmos de Classificação Binária. Os dados importantes aqui a se analisar são Renda, Idade, Patrimônio, já foi inadimplente alguma vez, Profissão e movimentação financeira nos últimos tempos.