



Seleção de Características: Métodos para Identificar as Variáveis Mais Relevantes em um Conjunto de Dados

Lucas Guilherme Pires Martins¹

Marcelo da Silva dos Santos²

A seleção de características é um processo fundamental na análise de dados, com o objetivo de identificar as variáveis mais relevantes para um determinado problema. Esse processo permite reduzir a dimensionalidade dos dados, melhorar a eficiência computacional e aumentar a precisão dos modelos de aprendizado de máquina. Neste texto, abordaremos a importância da seleção de características, discutiremos alguns métodos utilizados atualmente e suas aplicações. A seleção de características envolve a identificação das variáveis mais informativas em um conjunto de dados, descartando as variáveis menos relevantes ou redundantes. Existem diversos métodos disponíveis, cada um com suas próprias abordagens e critérios de avaliação. Entre os métodos mais utilizados atualmente estão: a análise de variância (ANOVA), a correlação de Pearson, a regressão logística e os algoritmos baseados em árvores de decisão, como o Random Forest e o Gradient Boosting. A ANOVA é amplamente aplicada quando há interesse em verificar a diferença estatisticamente significativa entre grupos de dados. A correlação de Pearson mede o grau de associação linear entre duas variáveis, sendo útil para identificar relações lineares em um conjunto de dados. Já a regressão logística é utilizada para análise de dados categóricos, permitindo estimar a probabilidade de ocorrência de um evento com base nas variáveis independentes. Os algoritmos baseados em árvores de decisão têm ganhado destaque devido à sua capacidade de lidar com dados complexos e não lineares. Eles fornecem medidas de importância das características, como a importância de Gini ou o ganho de informação, que auxiliam na seleção das variáveis mais relevantes. A seleção de características desempenha um papel crucial na análise de dados e no desenvolvimento de modelos preditivos. Ela permite reduzir a dimensionalidade dos dados, melhorar a eficiência

¹ Graduando em Sistemas para Internet. Faculdade QI Brasil. lucaspres567@gmail.com

² Mestre em Computação Aplicada. Faculdade QI Brasil. marcelo.santos@qi.edu.br

FAQI

JORNADA ACADÊMICA

12/06 A 14/06

<https://innova.faqi.edu.br/index.php/jornada>



computacional e aumentar a precisão dos modelos. Os métodos mencionados neste texto são apenas alguns exemplos dos muitos disponíveis na literatura científica. A escolha do método mais adequado depende do tipo de dados, do problema em questão e dos objetivos da análise. Em suma, a seleção de características é uma ferramenta indispensável na área de aprendizado de máquina, contribuindo para a obtenção de resultados mais precisos e relevantes.

Palavras chaves: Seleção de características, variáveis mais relevantes, métodos, análise de dados, aprendizado de máquina.