

Exercício Computacional 1

Tarefa: Classificação de Dados

A Câmara dos Deputados oferece um serviço de disponibilização de Dados Abertos. Um das fontes de dados é sobre gastos da Cota para Exercício da Atividade Parlamentar.

“Este serviço de Dados Abertos disponibiliza arquivos XML, JSON, CSV e XLSX compactados para download contendo os dados relativos aos gastos parlamentares registrados na Câmara dos Deputados. O acesso é gratuito e não requer autorização de uso. A intenção do serviço é permitir que os dados possam ser importados por sistemas de gerenciamento de banco de dados (SGBD) externos à Câmara dos Deputados para serem manipulados de acordo com as necessidades dos diversos setores da sociedade. Os arquivos são atualizados diariamente.” (Câmara dos Deputados, acesso em 15/04/17, <http://www2.camara.leg.br/transparencia/cota-para-exercicio-da-atividade-parlamentar/dados-abertos-cota-parlamentar>)

Pode-se acessar também “as explicações pertinentes a cada elemento de dado publicado como "Dados Abertos" acerca das despesas realizadas e publicadas na Transparência, quanto ao uso da CEAP (Cota para o Exercício da Atividade Parlamentar). Essas despesas se confirmam por meio de lançamento de débitos contra a cota do deputado e podem decorrer em razão de uma das causas a seguir: pelo reembolso de documentos fiscais emitidos (Notas Fiscais, Recibos ou Despesa no Exterior); ou pelo custeio de despesas telefônicas; ou pelo débito de requisição de serviços postais; ou pela emissão de bilhetes aéreos.” (Câmara dos Deputados, acesso em 15/04/17, <http://www2.camara.leg.br/transparencia/cota-para-exercicio-da-atividade-parlamentar/explicacoes-sobre-o-formato-dos-arquivos-xml>)

Uma seleção de dados de gastos da Cota Parlamentar de 2016 está disponível no site da disciplina para sua investigação. Nesta base foram resumidos os gastos mensais de cotas dos deputados em algumas categorias de despesas. A cada mês foram feitas a soma e a contagem de gastos nestas categorias, havendo dados ausentes de soma nos meses em que não houveram gastos na categoria.

Sua tarefa é analisar e selecionar os dados (atributos e instâncias), pré-processar os dados, extrair modelos de classificação e analisar os resultados. A variável de classe deve ser obtida a partir da discretização de um dos atributos de Soma de Valores de Documento (SomaVlrDocumento), devendo você escolher como discretizar e depois escolher as classes de interesse dentre as obtidas.

Faça a descrição dos dados selecionados, escolha e aplique técnicas de pré-processamento, aplique os algoritmos de classificação kNN, Árvore de Decisão e Naive Bayes (ou ainda outros), escolha o método e medidas de avaliação. Considere variações no pré-processamento, no algoritmo (parâmetros, escolhas) e no método de avaliação. A avaliação do modelo de classificação deve considerar que este modelo deve fornecer uma predição adequada para dados não vistos.

Elabore um relatório (de 4 a 6 páginas, formato SBC) descrevendo e justificando seu procedimento (metodologia) experimental, incluindo nome de programas e ferramentas utilizadas, apresente os resultados (gráficos e/ou tabelas) e faça discussão dos resultados encontrados.

Seu relatório deve ser entregue, por email (aulas+pgca@artificial.eng.br) e será apresentado em sala de aula.