

Descrição do projeto prático da disciplina Aprendizado Descritivo 2025/1

Renato Vimieiro

2025-06-02

Objetivos

O principal objetivo do projeto é explorar de forma prática os conceitos vistos na disciplina. Os alunos se reunirão em grupos e simularão o processo de análise de dados através de aprendizado descritivo. Para isso, deverão obter os dados, limpá-los/pré-processá-los, e aplicar alguma(s) técnica(s) de aprendizado descritivo vista no curso. Os grupos também deverão analisar os resultados e comunicá-los de forma escrita e oral. Dessa forma, o aluno irá experienciar os desafios relacionados a análise de dados em ambientes mais reais. A apresentação escrita e oral dos resultados servirá também como experiência para o grupo tanto do ponto de vista acadêmico, na apresentação de resultados em forma de artigo e palestras públicas, quanto do ponto de vista de formulação de relatórios que possam ser absorvidos por pessoas não diretamente relacionadas ao problema investigado.

Tarefas

São tarefas do projeto:

1. Obter dados para executar a análise. Os dados têm de ser representativos de problemas reais. Ou seja, dados obtidos de repositórios de benchmarks para Aprendizado de Máquina, tais como UCI ML Repository, não constituem dados viáveis para o projeto. Outras fontes como o Kaggle podem ser consultadas, porém, deve-se ter cuidado com o conjunto de dados escolhido. Conjuntos muito simples podem não permitir a aplicação das técnicas.
2. Limpeza dos dados. Tendo obtido os dados, os alunos deverão aplicar as transformações necessárias para tratar os dados e adequá-los à análise que será aplicada. Entre as possíveis tarefas podem estar eliminação/preenchimento de dados ausentes, mudança de formato dos dados, codificação ...

3. Aplicação de aprendizado descritivo. Nessa etapa, os alunos deverão aplicar alguma(s) técnica(s) vista(s) no curso para analisar os dados e extrair padrões interessantes dos mesmos. Podem ser aplicadas tanto técnicas supervisionadas quanto não-supervisionadas. **Recomenda-se que os alunos não implementem as técnicas, e tentem obter os algoritmos em algum repositório ou pacote.** Existem diversos pacotes para diversos ambientes com praticamente todos os algoritmos vistos nas aulas.
4. Análise dos resultados. Após a aplicação dos métodos, os alunos devem explorar os resultados, interpretando-os em busca de padrões interessantes. Deve-se levar em consideração aqui as métricas de qualidade dos padrões.
5. Documentação/apresentação dos resultados. Os alunos devem narrar/documentar o processo que executaram durante a análise, bem como os resultados obtidos. Essa documentação se dará de forma escrita através de **artigo científico**, e de forma oral através de apresentação em sala de aula virtual. O artigo terá no máximo 12 páginas em formato da SBC. A apresentação terá no máximo 15min, e mais 5min para discussão com a turma.

O que entregar

Os alunos devem criar um repositório público GitHub contendo todo o material usado nas análises (pacotes de software não devem ser incluídos no projeto, e sim referenciados para permitir a reprodução). Os dados também não precisam ser compartilhados. No entanto, caso sejam abertos, colocar o link para o local onde possam ser obtidos.

No Moodle, devem ser entregues: o artigo em pdf, os slides, e o link para o repositório com o conteúdo das análises.

Organização

O projeto pode ser executado em grupo de até 5 integrantes. Para diminuir a carga de trabalho de todos os envolvidos, é recomendável que o projeto seja desenvolvido em grupo. Além disso, cada grupo tem que designar atividades a todos os seus membros. Devem ser evitadas divisões de tarefas interdependentes. O ideal é que todos estejam trabalhando ao mesmo tempo.

Os grupos devem eleger um representante para enviar mensagem ao professor indicando o nome completo e matrícula de todos os membros. Essa tarefa é importante para que o professor possa dimensionar o número de aulas necessárias para apresentação dos projetos. Ela deve ser realizada o mais breve possível, até o dia 15/06.

Datas

A entrega do projeto será no dia **03/07**. As apresentações serão nos dias 03, 08 e 10/07.