

Universidade Estadual de Feira de Santana
Engenharia de Computação - EXA864 Mineração de Dados
Prof. Rodrigo Tripodi Calumby

Projeto Prático 2
“There and back again”

Objetivo Geral

Nesta atividade, a equipe deverá aplicar técnicas de mineração de dados para integrar, analisar e prever informações no contexto de famosos personagens de ficção e seus super-poderes. A atividade a ser desenvolvida tem um objetivo principal que é uma avaliação experimental rigorosa da predição de características a partir dos atributos dos personagens. Para tanto, deve-se considerar:

1. Definição e realização de procedimentos experimentais para construção de preditores eficazes.
2. Avaliação de múltiplos algoritmos de predição (pelo menos 2).
3. Avaliação de modelos de ensemble learning (pelo menos 1).

Objetivos Específicos

1. Realizar a integração das bases de dados.
2. A equipe deve realizar o pré-processamento necessário dos dados para torná-los adequados ao processo de mineração (limpeza, tratamento de dados ausentes, padronização, amostragem, conversão e seleção de atributos, etc.);
3. Definir quais algoritmos serão utilizados/desenvolvidos;
4. Definir a metodologia experimental de execução/validação das propostas (configurações dos algoritmos, protocolos experimentais, medidas de eficácia, etc.);
5. Definir os modelos a serem utilizados para visualização e análise de resultados.
6. Analisar os resultados e discutir as soluções propostas e desafios enfrentados.

Avaliação Experimental

Nesta atividade, a equipe deverá realizar uma avaliação experimental de algoritmos de predição de características ou super-poderes. Em um cenário real, este processo corresponderia, e.g., a prever se um determinado personagem possui ou não um determinado super-poder, considerando suas demais informações (incluindo suas características gerais e seus demais super-poderes). Um processo experimental deverá ser realizado para estimar a qualidade do classificador desenvolvido e garantir a construção de preditores com a maior qualidade possível e estimativas de eficácia confiáveis. Os resultados alcançados devem ser discutidos e as conclusões apresentadas. Deverão ser construídos preditores para: *Flight*, *Super Strength*, *Accelerated Healing*, *Alignment*, *Publisher*, *Gender* e mais um outro escolhido pela equipe.

Produtos: Proposta experimental; Código fonte (ou workflows); Artigo; e Apresentação.

Regras Gerais

- Os dados a serem utilizados são aqueles fornecidos pelo professor.
- O projeto deverá ser desenvolvido em equipes de três pessoas.
- O estudo experimental alvo desta atividade poderá ser realizado com auxílio de ferramentas gráficas ou via linguagem de programação e respectivas bibliotecas/frameworks.
- O código fonte deverá ser devidamente documentado.

Produtos

- Proposta experimental
- Códigos, *scripts*, *workflows* e dados que forem utilizados/desenvolvidos no projeto.
- Artigo Científico.
- Apresentação.

Proposta Experimental

- 1 página (fonte Times New Roman tamanho 10, espaçamento simples e margens 1,5cm)
- Deve descrever a metodologia a ser utilizada: algoritmos, medidas, protocolos, etc.

Artigo

- O artigo deve ser produzido seguindo o [modelo de formatação IEEE Transactions](#). Utilize preferencialmente o *template* LaTeX. O artigo deve ter no máximo 6 páginas (incluindo imagens e referências) e deverá ser entregue em formato PDF.
- Estrutura sugerida:
 - Introdução
 - Análise Proposta e Metodologia
 - Resultados e Discussão
 - Conclusões
 - Referências
- A estrutura do artigo pode ser adaptada de acordo com as necessidades da equipe.

Entrega

- A Proposta Experimental deve ser entregue (impressa ou via e-mail) no dia **13/08** para discussão junto com o professor.
- A entrega do projeto (artigo e código fonte) deverá ser realizada **até as 18h do dia 03/09/2019**.
- As apresentações ocorrerão nos dias **05/09 e 10/09** em ordem a ser sorteada no início da aula.
- Os produtos devem ser enviados em um único arquivo compactado (formato zip) via *Google Sala de Aula*.
- Os slides da apresentação (em formato PDF) deverão ser publicados no *Google Sala de Aula* logo após apresentação.

Apresentação

- Duração de 25 minutos (+ 5 minutos para perguntas).
- Não haverá tolerância de tempo adicional.
- Todos os membros da equipe deverão participar da apresentação.

Critérios Gerais de Avaliação

- Atendimento ao prazo de entrega estabelecido.
- Completude do material entregue.
- Atendimento aos objetivos, requisitos e orientações desta proposta.
- Adequação do protocolo experimental aplicado aos dados.
- Amplitude dos experimentos e rigor científico.
- Qualidade da discussão dos resultados e conclusões.
- Documentação do código fonte.
- Qualidade da apresentação.
- Artigo:
 - Atendimento à formatação indicada.
 - Clareza e objetividade.
 - Atendimento às normas de ortografia e gramática.
 - Adequação das citações.
 - Organização das informações.

Observações

- Fraudes serão penalizadas com nota 0.0 (zero) na atividade.
- Entregas com atraso de até 24h serão penalizadas em 2 (dois) pontos. Após 24h, entregas não serão mais aceitas, a menos que seja comprovado motivo de força maior que comprometa toda a equipe.
- Os slides das apresentações de todas as equipes serão coletados no início da aula.