



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
Programa de Pós-Graduação em Ciência da Computação
CMP259 – Descoberta de Conhecimento em Bases de Dados
Profa. Karin Becker

Desenvolver uma aplicação do seu interesse de descoberta de conhecimento (KDD) a partir de dados (aka. Ciência de dados), para um problema de seu interesse. Espera-se que o resultado do trabalho relate uma experiência a um problema válido, com seus resultados e lições aprendidas.

Objetivos:

- desenvolver as etapas de um processo KDD, de acordo com um objetivo de negócio;
- definir um objetivo de mineração alinhado com o objetivo de negócio e dados compatíveis;
- ser capaz de identificar dados adequados aos objetivos definidos, compreender suas características e avaliar sua sanidade;
- ser capaz de escolher técnicas de mineração de dados adequadas, e de aplicá-las (incluindo preparação de dados, escolha do algoritmo, parametrização, interpretação de resultados);
- ser capaz de escolher e utilizar ferramentas adequadas às diferentes tarefas do processo de KDD;
- ser capaz de enquadrar a experiência em um processo KDD, e ter um posicionamento crítico quanto às escolhas feitas, e seus resultados.
- ser capaz de relatar os resultados e sua contribuição através do problema em um artigo científico.

Quem:

Recomenda-se o desenvolvimento do trabalho em duplas. Pode ser feito individualmente.

Opções de trabalho:

- Trabalho de desenvolvimento:** objetivo é passar por todas as etapas do processo de maneira crítica.
- Artigo de pesquisa,** representando um estágio parcial de sua pesquisa e que resulte em uma publicação científica potencial: o objetivo é apresentar um projeto de pesquisa onde a aplicação de técnicas de extração de conhecimento agregou valor ao problema. Existem muitos eventos com datas próximas à da disciplina: pesquise, e discuta com a professora algumas opções !

Cronograma Atualizado:

Etapa	Entregável	Entrega	Peso
0 – Informar dupla	a) Nome dupla	15/10	
1 – Compreensão do Negócio e dos Dados	b) Relatório c) Apresentação (max 12')	a) 27/10 b) 31/10	20%
2 – Mentoria e Discussão de resultados preliminares		28/11 e 5/12	10%
3 – Trabalho Final	a) Relatório/Artigo b) Apresentação (15'-20')	12/12 ou 19/12 (relatório no dia da apresentação)	70% (35% + 35%)

Etapa 1: Proposta de Trabalho

Nesta etapa, devem ser apresentados os resultados das fases de Compreensão do Negócio e dos Dados para um problema de relevância. Isto inclui :

- a definição de um *objetivo de negócios* (objetivo primário e, se pertinente, objetivo secundário)
- a definição de um ou mais *objetivos de mineração* de permitam a resposta a estes problemas
- a busca de *dados compatíveis* com os objetivos definidos;
- detalhes sobre o processo de *extração dos dados* (fontes originais, forma de obtenção)
- a exploração *preliminar* destes dados, quer na forma de estatísticas ou exploração visual (se pertinente)
- definição de um plano preliminar para a execução das próximas etapas do processo
- *recursos* utilizados.

Entregáveis (Etapa I):

- Relatório em formato livre, de no máximo 5 paginas, que contenha os itens solicitados.
- Apresentação: descrevendo o essencial da proposta de trabalho

Etapa 2: Mentoria (10%)

- Discussão dos resultados preliminares em reuniões privadas com a professora, **a marcar**. Avaliação é feita com base nos avanços realizados. Agendamentos FIFO.

Entregáveis (Etapa II): material mostrado na reunião.

Etapa 3: Nesta etapa, devem ser apresentados os resultados obtidos com seu projeto. Espera-se que tenham ocorrido todos os realinhamentos necessários entre objetivos (de negócio/mineração) e resultados obtidos.

Material a ser entregue (Etapa III):

- a) **Abordagem Desenvolvimento.** Relatório de no máximo 12 páginas. A estrutura e o formato são livres, e através do conteúdo, busca-se poder avaliar os seguintes pontos:
 - Contexto, problema de negócio e objetivo, e justificativa do valor agregado do conhecimento a ser obtido;
 - Objeto ou tarefa de mineração de dados (quais padrões são buscados que podem resultar em conhecimento);
 - Dados a serem utilizados, profiling, síntese da exploração (o que for relevante);
 - detalhes relevantes do pré-processamento realizado;
 - detalhes da mineração de dados (incluindo justificativa de escolha dos algoritmos, e detalhamento dos parâmetros)
 - análise dos resultados obtidos com avaliação crítica.
 - Ciclos, se existirem.
 - recursos utilizados (total liberdade é dada para a escolha das ferramentas de apoio, podendo ser uma combinação de ferramentas).
 - lições aprendidas
- b) **Abordagem artigo científico:** escolha uma conferência alvo, e use o formato e tamanho de páginas alvo, com duas páginas extras de bônus. As referências não contam no tamanho máximo. Sugestão de estrutura: além do título e resumo (abstract), deve conter: 1)

Introdução; 2) Trabalhos Relacionados; 3) Descrição da Proposta (ou Materiais e Métodos); 4) Experimentos (ou Avaliação); 5) Ameaças de validade; 6) Conclusões e Trabalhos futuros. Para a introdução, sugere-se o formato MIT (contexto, problema é difícil, concorrentes não resolveram e tem um gap, objetivo, contribuição, estrutura do artigo). No artigo, seu objetivo é mostrar sua contribuição ao problema de pesquisa através da extração de padrões dados, em uma abordagem data-driven.

- Apresentação dos aspectos principais em aproximadamente 15-20 minutos.