

Especificação do Projeto Final

CDIA | 2022

Introdução

Este documento especifica o projeto que cada grupo deverá desenvolver como produto das disciplinas: Projeto Integrado, Banco de Dados e Engenharia de Software.

Descrição do Projeto

A partir de um objetivo de negócio escolhido por cada grupo, o projeto consiste em sistematizar as técnicas apresentadas durante o semestre para atingir os objetivos.

Método

Cada projeto deve seguir o seguinte método:

1. Oferecer uma contextualização do contexto do negócio e da necessidade da solução que seu grupo irá propor.
2. Estabelecer claramente o objetivo do trabalho
3. Implementar o projeto
4. Analisar como a implementação atende ao objetivo proposto
5. Conclusão

Composição dos grupos

Cada grupo deve ter entre 2 e 6 alun@s. Trabalhos individuais não são permitidos.

Requisitos do projeto

Os projetos devem utilizar as mesmas técnicas apresentadas durante o curso para cumprir os requisitos.

Banco de Dados

Cada projeto deve considerar os seguintes requisitos:

- REQ#01: Definir pelo menos uma função para realizar tarefas específicas
- REQ#02: Identificar um dataset (não pode ser *toy*) que sobre a temática do projeto
- REQ#03: Construir um modelo conceitual
- REQ#04: Construir um modelo lógico
- REQ#05: Construir um físico

- REQ#06: Popular o BD a partir do dataset
- REQ#07: Criar 10 questões para que o BD responda
- REQ#08: O relatório do projeto deve ser desenvolvido e entregue em um caderno Jupyter.
- REQ#09: O projeto deve ser apresentado para a banca na data estipulada.

Engenharia de Software

- REQ#01: Propor ao menos cinco variáveis de estado com os seus respectivos domínios de valores no mundo lógico.
- REQ#02: Especificar um predicado de estado inicial, uma função de estado, um predicado de estado e uma ação, pelo menos, envolvendo as variáveis de estado propostas no mundo lógico.
- REQ#03: Ilustrar os predicados, funções e ações no contexto de estados específicos do negócio do projeto.
- REQ#04: Implementar as variáveis de estado, os predicados, funções e ações em Python usando a arquitetura de referência (módulos de entrada, saída, controle, modelo e validação).
- REQ#05: Mostrar que o código em Python produz os resultados especificados no mundo lógico.
- REQ#06: Documentar o modelo lógico e o de implementação na linguagem Python em um caderno Jupyter.

Machine Learning

- REQ#01: Utilizar um ou mais datasets (não pode ser *toy*) para o treinamento dos classificadores (*)
- REQ#02: Realizar uma análise exploratória do dataset por meio de um caderno Jupyter. Utilize gráficos na análise.
- REQ#03: Treinar um classificador k NN
- REQ#04: Treinar um classificador de *Regressão Linear*
- REQ#05: Treinar um classificador de *Regressão Logística*
- REQ#06: Treinar um classificador *Naive Bayes* (Multinomial, Bernoulli **ou** Gaussian)
- REQ#07: Treinar um classificador *Support Vector Machine* (SVM) (*desejável*)
- REQ#08: Treinar um classificador de *Árvore de Decisão* ou *Floresta Aleatória*
- REQ#09: Todos os classificadores devem ser avaliados com 30% dos dados, utilizando as métricas F1, acurácia, revocação (*recall*), precisão. Além disso, utilizem a matriz de

confusão para a visualização do desempenho

- REQ#10: Todos os classificadores devem ser persistidos (*joblib, pickle*) antes de serem entregues
- REQ#11: O projeto-final de contar com um modelo por integrante do grupo. Isto é, um grupo com 2 integrantes pode escolher dois dos REQs #03 - #08. Um grupo com 6 integrantes deverá treinar todos.

(*) Para atender todos os requisitos do projeto, vocês podem complementar o dataset com dados realistas, assim como feito nos cadernos e no projeto-piloto.

Como cada projeto será avaliado?

Para cada disciplina, a nota final do projeto (A_{proj}) será dada da seguinte forma:

R = cumprimento dos requisitos (0 = baixo; 5 = médio; 10 = alto)

M = aplicação do método (0 = baixo; 5 = médio; 10 = alto)

O = relevância do objetivo de negócio e do trabalho (0 = baixo; 5 = médio; 10 = alto)

J = qualidade da apresentação do relatório Jupyter (0 = baixo; 5 = médio; 10 = alto)

B = qualidade da apresentação (powerpoint e oral) para a banca (0 = baixo; 5 = médio; 10 = alto)

$$A_{proj} = \frac{(R + M + O + J + B)}{50}$$

Como cada alun@ será avaliada@?

Cada membro do grupo fará uma autoavaliação e uma avaliação anônima da participação dos outros membros do grupo na elaboração do projeto. Em todas as avaliações, será necessário justificar a nota atribuída a si e aos pares. Essa avaliação é conhecida como avaliação 360. A nota a_{360} de cada alun@ será dada pela média entre as avaliações dos pares e a autoavaliação.

A nota final de projeto de cada alun@ é dada por:

$$A_{est} = A_{proj} \cdot \frac{a_{360}}{10}$$

Prazos

Acompanhamento

A partir de 01/06, todos os grupos devem reportar o progresso do trabalho **semanalmente**.

Prazo final

O trabalho deve ser entregue até o dia 17/06 (**impreterivelmente**).

