

Enunciado Trabalho T1

O trabalho T1 consiste em três etapas vão desde a implementação até a aplicação de dois métodos de classificação vistos em aula: kNN e Naïve Bayes. Abaixo a descrição de cada uma das etapas do trabalho e sua pontuação respectiva.

Etapas 1 - Implementação (70% da nota):

Nessa etapa deverá ser feita a codificação dos algoritmos kNN e Naïve Bayes vistos em aula, os requisitos para cada uma das implementações seguem abaixo.

kNN:

- Algoritmo para busca do vizinho mais próximo: Força Bruta e KD Tree;
- Métrica: Distância Euclidiana
- k: 1, 3 ou 5 para força bruta, k=1 para KD Tree
- Tipos de Features: Numéricas

(Gaussian) Naïve Bayes:

- Tipos de Features: Numéricas (Usar uma distribuição gaussiana/normal para aproximar os dados).

Os dois algoritmos devem ser de codificação própria do grupo, **não** usem bibliotecas prontas ou código copiado. A linguagem de programação é livre (na dúvida usem Python!).

Etapas 2: Treinamento e Inferência (30% da nota)

Ambos os algoritmos devem ser utilizados no *penguins* dataset disponibilizado no moodle. Deve ser realizado split entre treino e teste (70% Treino - 30% Teste). Os resultados devem ser mostrados em tabela de confusão. Computar os valores de acurácia, precisão (por classe), recall (por classe) e f1.

Para essa etapa do trabalho está liberado uso de bibliotecas (sklearn e outras).

Número de Integrantes: De 1 a 3 integrantes

Data de Entrega: 20/04/21

Formato da Entrega: Link no Moodle

Entregáveis:

Um arquivo .zip com o seguinte conteúdo:

- a. Arquivo TXT com Nome e matrícula de todos os integrantes do grupo;

- b. Arquivo README.txt com descrição dos demais arquivos no zip e eventuais procedimentos necessários para que o professor replique os resultados;
- c. Código fonte da implementação dos classificadores (Etapa 1);
- d. Código fonte para importação do conjunto de dados, matriz de confusão e cálculo das métricas (Etapa 2);
- e. Documento PDF com os resultados obtidos na Etapa 2 (Não precisa de muita formatação, uma listagem das duas matrizes de confusão e das métricas é o suficiente);
- f. Arquivo csv/txt com as classificações realizadas na Etapa 3 para o conjunto de testes;

Sugestão de Estrutura de pastas para o arquivo zip:

- ✓ Etapa 1
- ✓ Etapa 2
- ✓ integrantes.txt
- ✓ README.txt

Caso o grupo opte por desenvolver em python, um Jupyter Notebook pode ser usado para entregar os itens “d” e “e” no mesmo arquivo, dispensando a entrega de um pdf separado.