**Requisitos da disciplina: Recuperação de Informação Baseada em Conteúdo**

Descrição: Considerando o data warehouse construído (conforme especificado nos requisitos

anteriores), o grupo deve:

- Descrever como armazenar as imagens (ou outros dados complexos) no data warehouse

projetado e implementado

- Armazenar os vetores de características no data warehouse mantido no PostgreSQL

- Criar consultas por similaridade range query e kNN (aceita-se o valor fixo para k, ex: os 5

mais similares) sobre o data warehouse que mantém os vetores de características

Observação: utilizar os vetores de características disponíveis na Internet.

Exemplos:

*https://www.kaggle.com/datasets*

*http://archive.ics.uci.edu/ml/index.php*

*https://registry.opendata.aws*

[*https://toolbox.google.com/datasetsearch*](https://toolbox.google.com/datasetsearch)

Arquivos solicitados para entrega. O relatório (item I) deve conter os vetores de características

que são armazenados no data warehouse, além das respostas das consultas range query e kNN

escritas em SQL, utilizando os mecanismos que o professor irá descrever durante as aulas.

Nosso DW contém a modelagem para análise de dados de plantas. Uma das tabelas possui os vetores de características já calculados a armazenados. Assim, podemos utilizar consultas por similaridade para comparar as imagens, por exemplo, usando cálculos de funções de distância.

Em uma das nossas dimensões temos nossos vetores de características armazenados:

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Text

Description automatically generated

Acima, realizamos a criação das funções de similaridade para os cálculos de distância. Escolhemos como ponto de comparação a planta *bluebell* de id = 1

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application

Description automatically generated

Realizando as consultas vemos que a consulta por KNN identificou as flores *buttercup* e *crocus* como sendo as mais similares quando comparadas a nossa flor consultada (*bluebell*). Abaixo, temos imagens de exemplo: *bluebell* e *crocus*, respectivamente.

 