

Projeto Final do módulo de dados não estruturados

Objetivo

- Fixar o conteúdo trabalhado em aula
- Aprofundar em algum problema específico no campo de dados não estruturados

Entrega

01/06

Logística

1 Projeto por Squad

Squad de 3 pessoas ou mais

Onde e como Entregar

Me envie o .ipynb ou .zip por email com nome do arquivo: 546-[squad]-[projeto].ipynb, para ph@retinavision.com.br

Critérios da Avaliação:

- Legibilidade e organização dos códigos e notebook, replicação.

-

Projeto de Imagem

Introdução

A fabricante de chocolate L&L, preocupada com a sua gestão de produção e como pode trazer inovação trouxe o desafio de melhorar o processo de contagem dos chocolates L&L's como mostrado na imagem. A ideia trazida foi de mantendo o mesmo maquinário utilizar uma câmera fixada na esteira de produção fazer a contagem de L&L's com foco em controle de qualidade.

Objetivo

Montar um pipeline para contagem e extração de features (cor e tamanho) da linha de produção da fabricante de chocolates, L&L

O que é para fazer

Notebook com pipeline que faz o pré processamento, contagem dos L&L se extrai as features de área e cor, para no final saber quantos L&L tem, quais cores e os tamanhos, para ver se não tem nenhuma falha de qualidade.

Dataset

https://drive.google.com/file/d/1YGRoMo3iL0BaxN8x_-dIDeakCbbIYmOP/view?usp=sharing



Projeto de Áudio

Introdução

A Letsound é um serviço de streaming de música aberto para qualquer produtor musical importar e compartilhar suas músicas na internet. Devido ao alto volume de uploads, ficou difícil fazer uma classificação manual dos gêneros que são utilizados para os algoritmos de recomendação, assim, a Letsound os contratou para montar, preparar o dataset para serem usados no processo de treinamento dos modelos.

Objetivo

Montar o pipeline para pré processamento dos arquivos de áudio que serão usados para treino e teste nos modelos de classificação de gênero musical. Não será trabalhado o modelo, e sim, somente o pré processamento

O que é para fazer

- 1) Baixar o mini dataset <http://opihi.cs.uvic.ca/sound/mini-genres.tar.bz2> que contém uma amostra de diferentes gêneros de música
- 2) Carregar os áudio com um limite de 10s por arquivo
- 3) Montar os scripts para extrair as seguintes features (hop_len e n_fft, podem manter padrão):
 - Espectrograma em escala logaritmica de todos os áudios (figura)
 - Espectrograma da escala de mel (MFCC) para n_mcc=13 (figura)
 - Chroma feature (figura)

Dataset

<http://opihi.cs.uvic.ca/sound/mini-genres.tar.bz2>

Projeto de Texto

Introdução

A LetsBuy é um e-commerce de produtos eletrônicos com foco em trazer os melhores produtos e experiência de compra. Recentemente com o seu alto crescimento está com dificuldades para conseguir tratar o alto volume de reviews nos seus produtos para os ajudarem a direcionar qual deve ser a tendência de seu estoque e quais problemas de suporte que possa ter. Focando muito em tecnologia, resolveu desenvolver internamente um classificador de sentimentos em cima dos reviews feitos pelos seus clientes com objetivo de ter melhor sua operação.

Objetivo

Montar o modelo para classificação de sentimentos conforme base de treino e teste, focado em obter a melhor métrica de acurácia.

O que é para fazer

Fazer a modelagem, com as devidas análises, se necessário, pré processamento, modelagens, e fine tuning de hiperparâmetros com objetivo de montar o modelo com melhor acurácia para a tarefa de classificação de sentimentos dos reviews.

Dataset

<https://drive.google.com/file/d/1bc7xrLwzy3J6eI7xMG6WfIRm2cFQHijb/view?usp=sharing>