

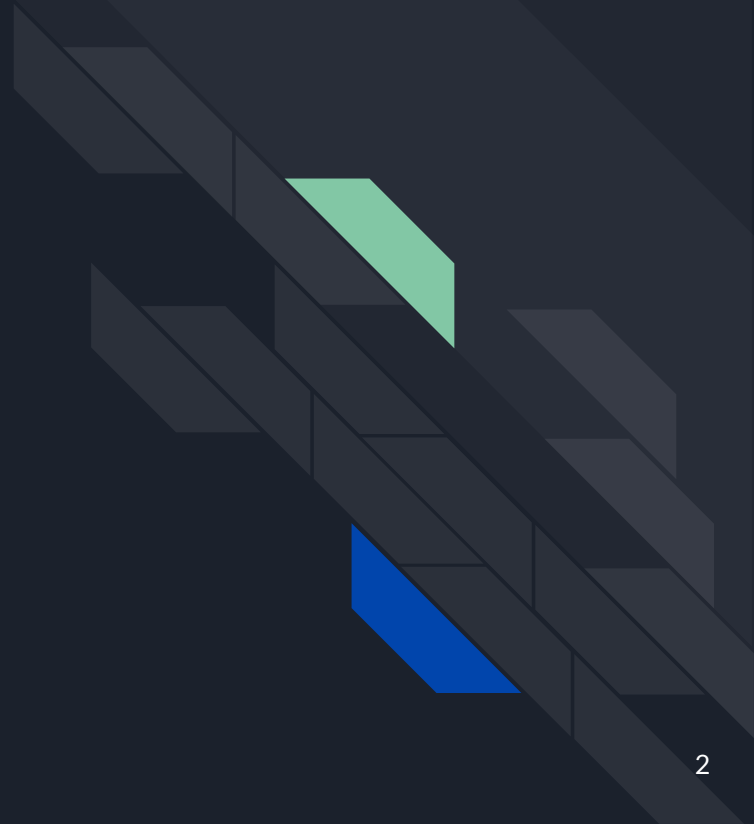


Classificação Hierárquica

Bruno Gomes Coelho - 9791160
Pedro Regattieri Rocha - 8531702



Tipos de Problema

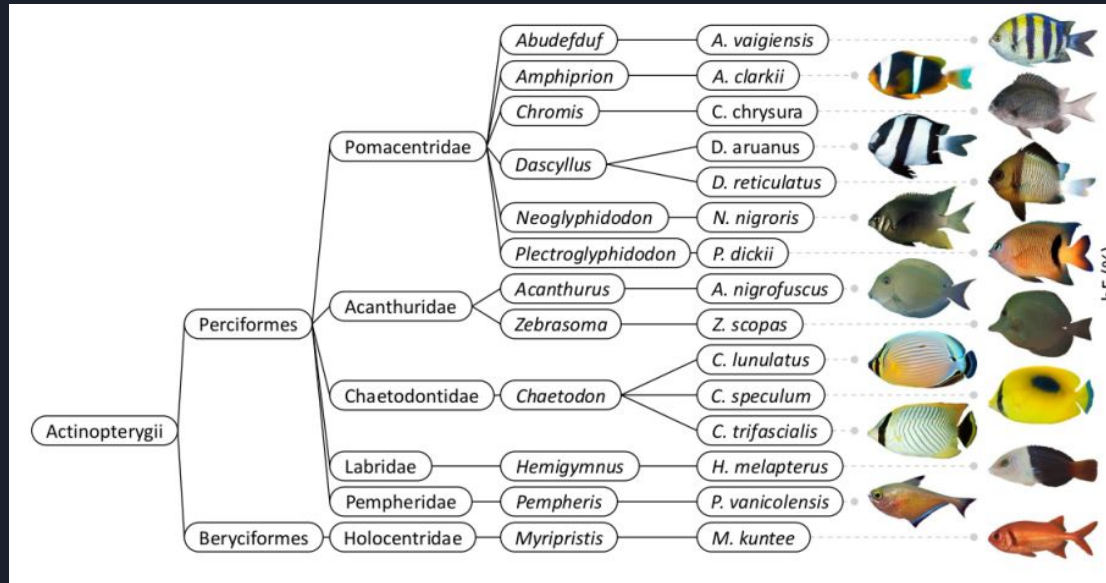




Hierarquia presente nas classes

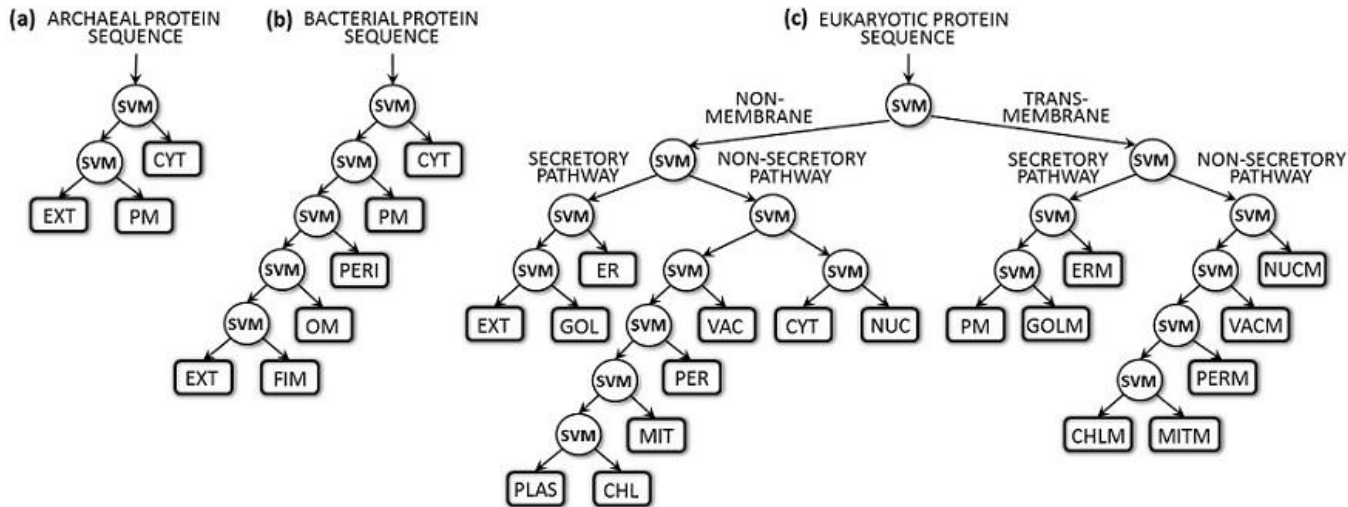
- Problema onde os dados existem com alguma **dependência** ou **taxonomia**.
- Existência de classes *superclasses* e *subclasses*.
- *Frequentemente usado na classificação de textos e seres vivos.*
- Idealmente utilizamos a estrutura do problema!
- Vamos ver alguns exemplos:

Classificação de Peixes



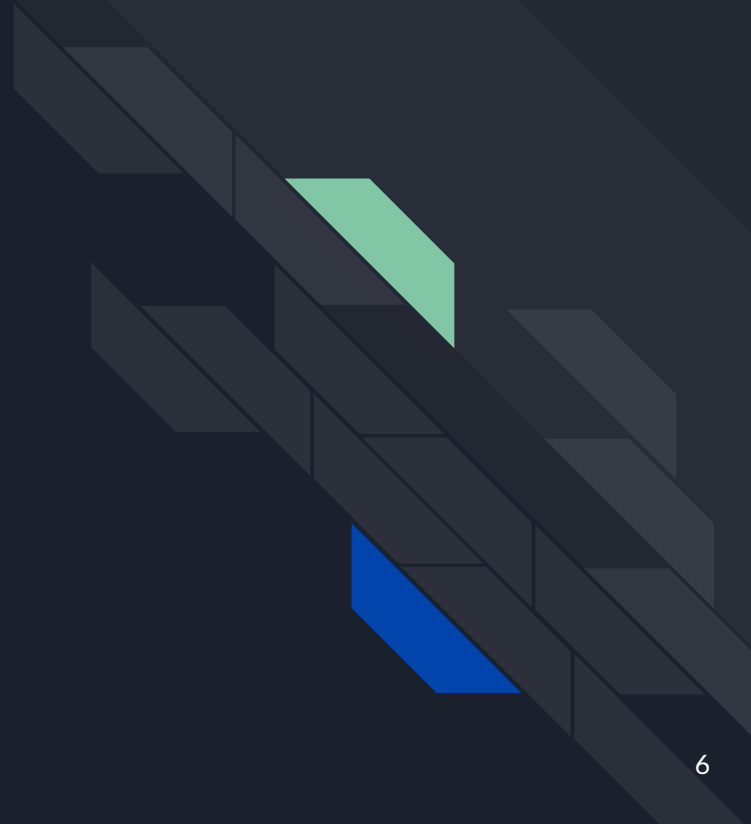
Parmezan, Antonio & Alves de Souza, Vinícius & Batista, Gustavo. (2018).
Towards Hierarchical Classification of Data Streams.

Localização da proteína



LocTree2 predicts localization for all domains of life
Bioinformatics. 2012 Sep 15;28(18):i458-i465.

Classificação





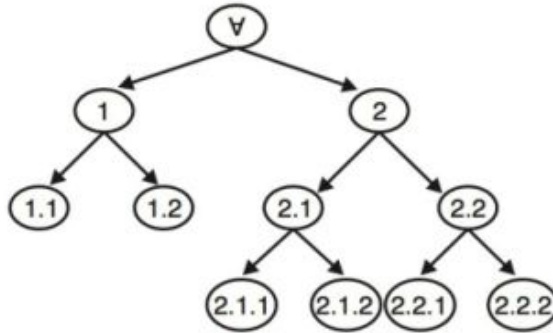
Classificação

- Tipo da hierarquia
- Caminhos dentro da hierarquia
- Profundidade

Tipo da hierarquia

Árvore

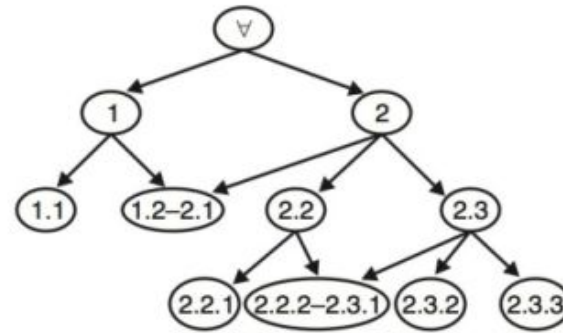
- Tecnicamente é um subtipo de DAG
- Todo nó possui apenas um nó pai



(a) Árvore

DAG

- “Direct Acyclic Graph”
- Pode ter mais de um nó pai



(b) DAG



Caminho

Único

- Em todo nó pai, só posso escolher 1 outro nó filho para seguir
- No final, terei no máximo 1 classificação de classe para cada instância

Multirrótulo

- Posso seguir mais de 1 caminho durante a classificação
- Isso equivale a um problema de classificação hierárquica multi rótulos
- Minha instância pode ser mais de uma coisa (eg: mochila de criança)



Profundidade

Parcial

- A classificação não é total - para em algum estado
- Comum em dados biológicos

Completa

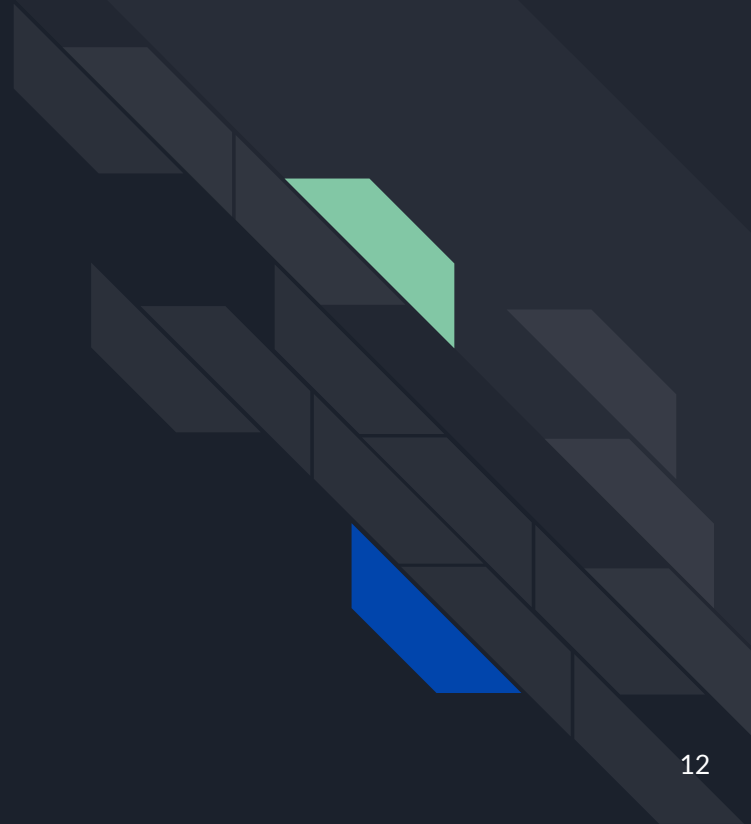
- Todo caminho resulta num nó folha



Na prática

- Problemas apenas multi rótulos não são geralmente considerados
- Mais comum:
 - Árvore
 - Apenas 1 camino
 - Profundidade completa

Algoritmos



Distinção de Algoritmos

- Para dividirmos os algoritmos de classificação hierárquica presentes na literatura, utilizamos quatro critérios principais (três dos quais estão diretamente relacionados à classificação do problema):
 1. Se o algoritmo é capaz de lidar com DAGs ou apenas com árvores (visto que, como árvores são tipos simplificados de DAGs, um algoritmo capaz de trabalhar em um DAG é capaz de funcionar em árvores, porém a recíproca não é verdadeira, ao menos sem o algoritmo antes ser adaptado para tal)
 2. Se o algoritmo é capaz de rotular os dados através de diversos caminhos ou não (ou seja, se o caminho é único ou multirrótulo).
 3. A profundidade das previsões do algoritmo, se ela é parcial ou completa
 4. Os classificadores utilizados para a resolução do problema: Globais ou Locais. Os locais são subsequentemente subdivididos em três tipos: um único classificador por nível, por nó pai, ou para cada nó.



Classificadores Locais

- Classificador por nó - Mais utilizado, implica na combinação de diversos classificadores binários nos níveis da hierarquia.
- Classificador por nó pai - Menos utilizado em DAG pois pode ocorrer redundância.
- Classificador por nível - O menos utilizado na literatura pois podem ocorrer conflitos entre os diferentes níveis.



Abordagem Top-Down

- Partindo dos níveis iniciais do DAG/Árvore, predições são utilizadas para escolher os próximos classificadores até um nó folha ser atingido (ou a execução do algoritmo ser interrompida por estar dentro de uma margem de confiança).
- Desvantagem: Um erro de classificação será propagado conforme a classificação continua.



Classificadores Globais

Big-Bang, One-Shot

- Utilizam a hierarquia durante a construção da classificação
- Técnicas mais complexas, que geram um algoritmo mais simples
- Mais interpretável
- Mais analisado para a classificação multi rótulos
- Estado da arte



C4.5-H

- Adaptação do clássico algoritmo de árvore de decisão C4.5
- Para classificação multi rótulos
- Cada nó da árvore prediz não 1 classe mas um vetor binário de tamanho C
- Utilizamos a entropia estendida para várias classes:

$$-\sum_{i=1}^N ((p(c_i) \log p(c_i)) + (q(c_i) \log q(c_i)))$$

- Capacidade reduzida



Clus-HMC

- Ainda classificação multi rótulos
- Adaptação da CART - pensando na ideia de **regressão**
- Definimos um vetor média de um conjunto a partir das classes
- Utilizando a distância euclidiana, calculamos a variância
- A variância será nossa medida da qualidade do corte
- Podemos ponderar pela distância na classificação - ideia de ancestralidade
- Utilizamos a média como vetor de classificação daquele nó



Comparação:

- Clus-HMC teve resultados estatisticamente melhores que o modelo top-down Clus e que o C4.5 H
- Overfittou menos devido a ter “apenas” 1 target
- Extremamente útil na interpretabilidade - aplicação direta em biologia

“Given that HMC decision trees can yield better efficiency and interpretability without suffering a decline in predictive accuracy compared to learning separate trees, their use should definitely be considered in HMC tasks where interpretable models are desired”



HMCN

- HMCN - Hierarchical Multi-Label Classification Networks
- Estado da arte - técnica desenvolvida nos últimos 5 anos
- Desenvolvido no Brasil, na UFSCar e no ICMC
- Parte do princípio: Como juntar as abordagens locais e globais?
- Utiliza os avanços recentes em redes neurais



HMCN

Motivação:

- A abordagem global aprende bem detalhes da hierarquia, mas tende a underfittar
- A abordagem local aprende bem regras locais, mas tende a overfittar.

Criar uma versão híbrida



HMCN

- 1 output por nível da hierarquia - abordagem local
- 1 output geral - abordagem global
- Cada loss local propaga os gradientes considerando o aprendizado daquele nível;
- A loss global mantém as dependências entre as hierarquias
- É acrescentado uma penalização por violar regras da hierarquia

HMCN

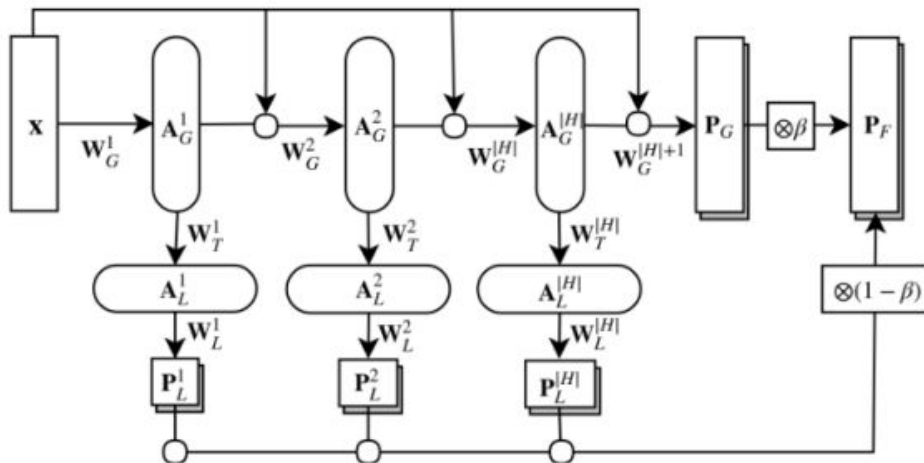


Figure 1. HMCN-F architecture.

$$\mathbf{A}_G^1 = \phi(\mathbf{W}_G^1 \mathbf{x} + \mathbf{b}_G^1)$$

$$\mathbf{A}_G^h = \phi(\mathbf{W}_G^h (\mathbf{A}_G^{h-1} \odot \mathbf{x}) + \mathbf{b}_G^h)$$

$$\mathbf{P}_F = \beta \left(\mathbf{P}_L^1 \odot \mathbf{P}_L^2 \odot \dots \odot \mathbf{P}_L^{|H|} \right) + (1 - \beta) \mathbf{P}_G$$



Outras abordagem

Classificação:

- Hierarchical Classification using a Competitive Neural Network (HC-CNN)
- Global-Model Naïve Bayes (GMNB)

Acredito que existe ainda muita inovação a ser feita na área



Avaliação

‘There seems to be no standards on how to evaluate hierarchical classification systems or even how to setup the experiments in a standard way.’

A survey of hierarchical classification across different application domains



Avaliação

- Reportar uma taxa de desempenho para o modelo como um todo;
- Reportar uma taxa de desempenho para cada nível da hierarquia separadamente;
- Reportar uma taxa de desempenho para cada classe.



Avaliação

- Em distância: Considera a quantidade de arestas
- Em profundidade: Pesa as arestas
- Em similaridade: Baseado em um vetor característico
- Relações de ancestralidade: número de nós em comum



Relações de ancestralidade

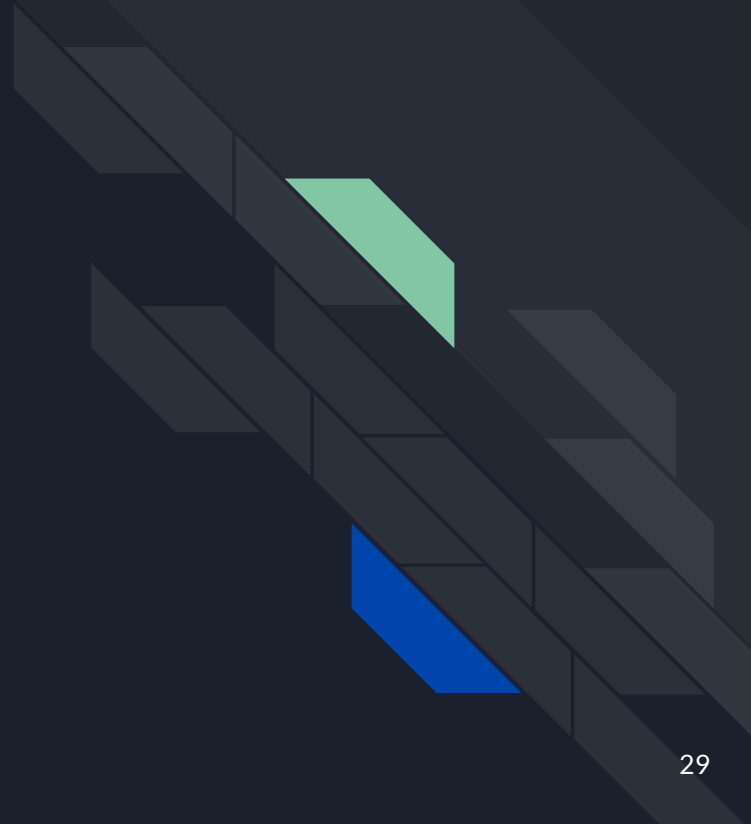
Precisão

Recall

$$P_H = \frac{|Ancestral(c_p) \cap Ancestral(c_v)|}{|Ancestral(c_p)|}$$

$$R_H = \frac{|Ancestral(c_p) \cap Ancestral(c_v)|}{|Ancestral(c_v)|}$$

Bases



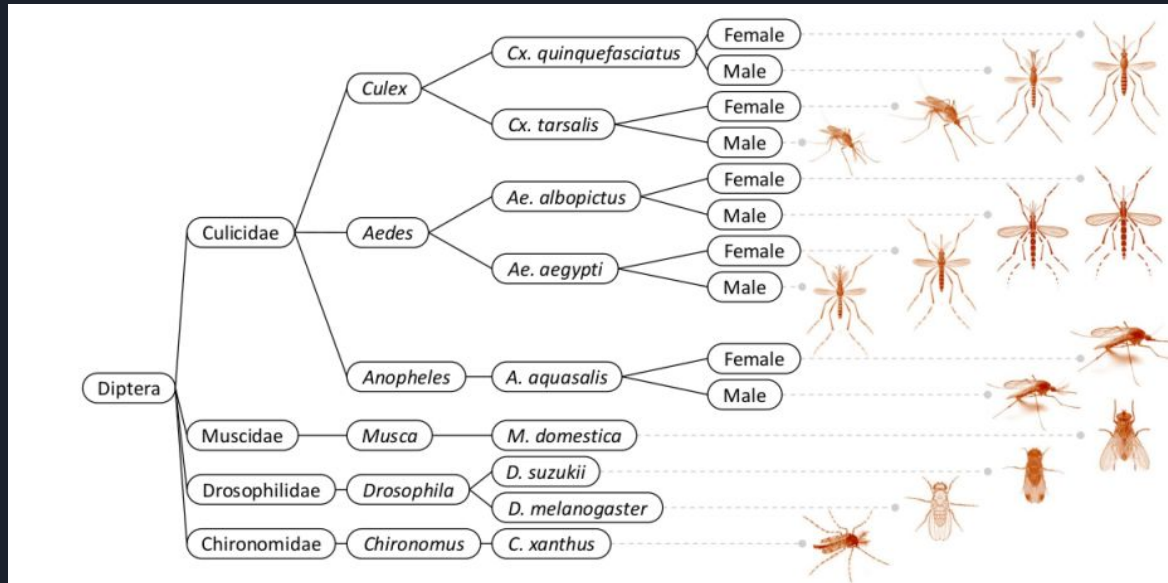


Classificação de Mosquitos

- Hierarquia de árvore, caminho único, profundidade total
- Motivação: Armadilha inteligente
- Classificação da espécie do mosquito a partir de seu batimento de onda
- 21,722 exemplos em 14 classes diferentes

Disponível [aqui](#).

Classificação de Mosquitos



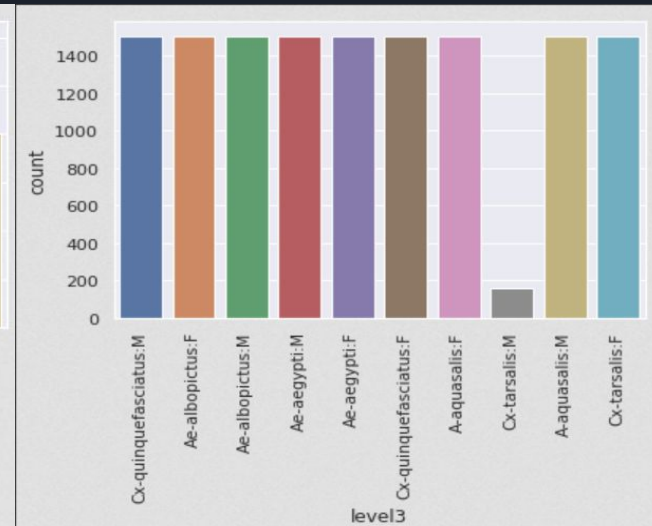
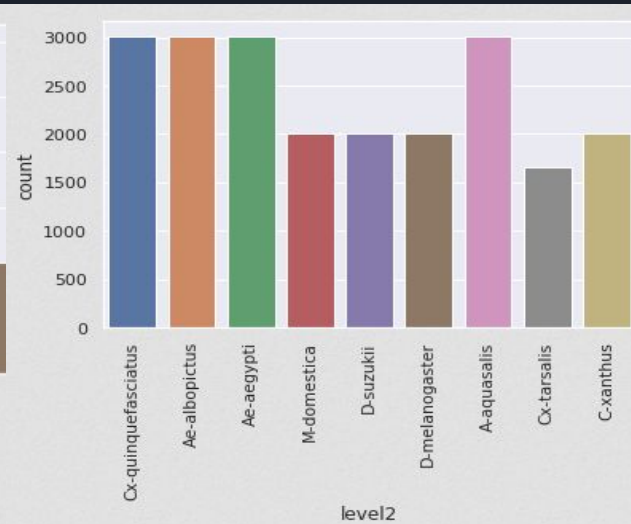
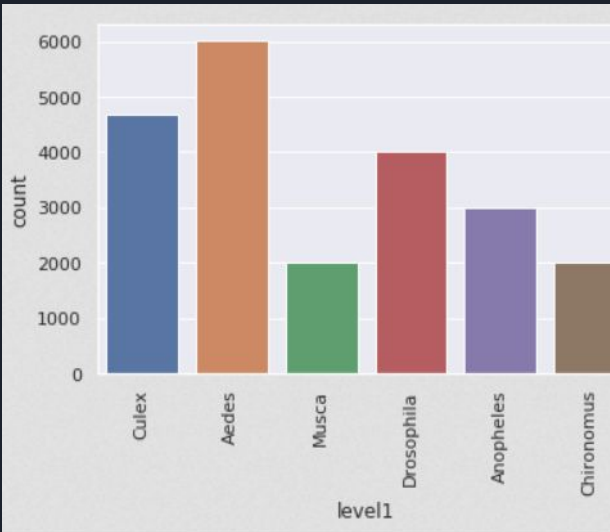
Parmezan, Antonio & Alves de Souza, Vinícius & Batista, Gustavo. (2018).
Towards Hierarchical Classification of Data Streams.



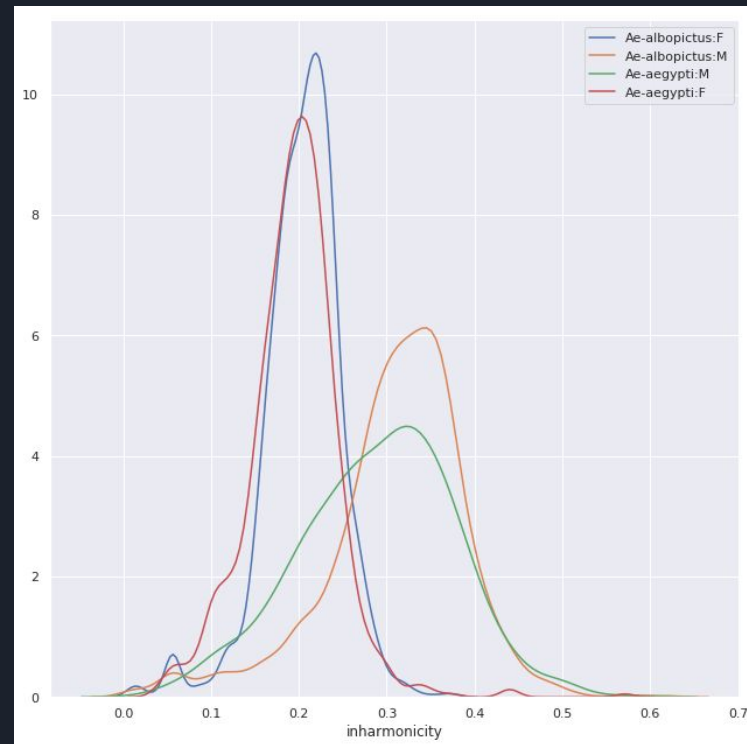
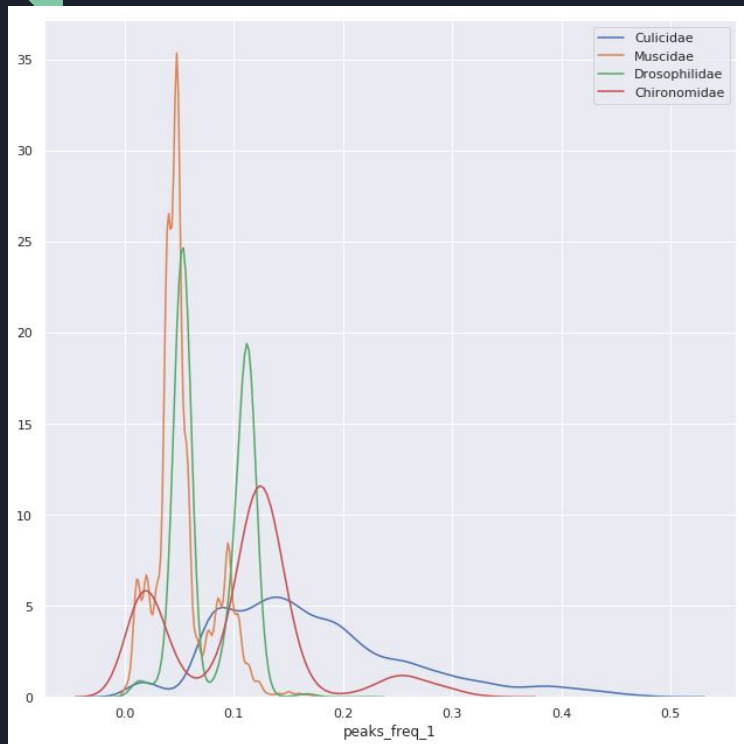
Classificação de Mosquitos

- Não possui nulos
- Extração de 33 features a partir do sinal de onda:
 - Frequência de picos
 - Transformada de Fourier
 - Inarmonicidade
- Divisão em sexo para algumas espécies
- 1 classe desbalanceada

Classificação de Mosquitos



Classificação de Mosquitos



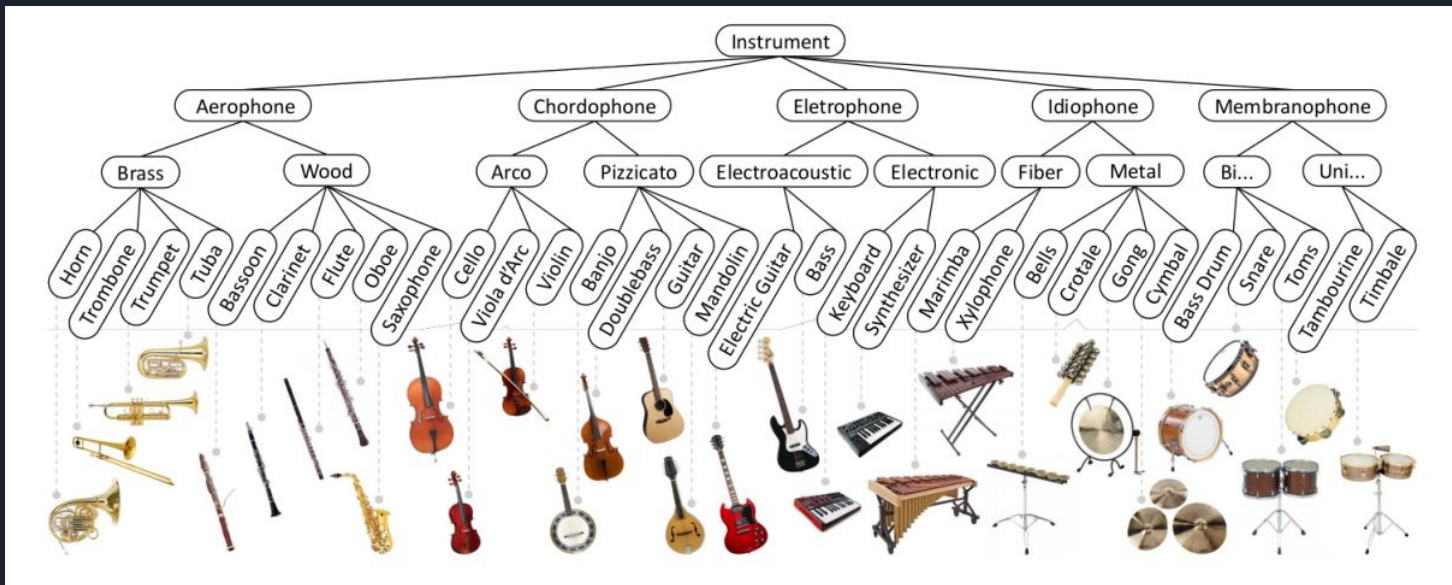


Classificação de Instrumentos

- Hierarquia de árvore, caminho único, profundidade total
- Classificar instrumentos musicais com base no sinal de áudio dele tocando
- Extração de 30 features - Mel Frequency Cepstral Coefficients
- 9,419 exemplo distribuídos em 31 classes

Disponível [aqui](#).

Classificação de Instrumentos



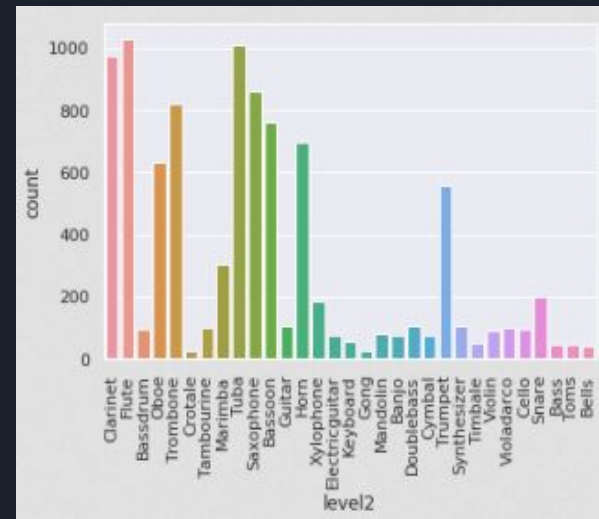
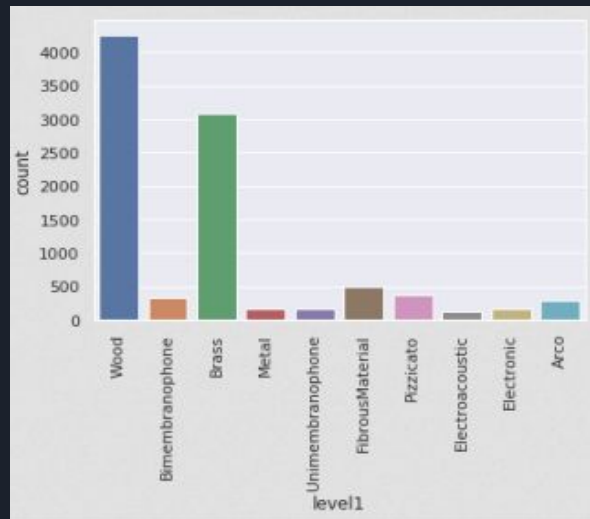
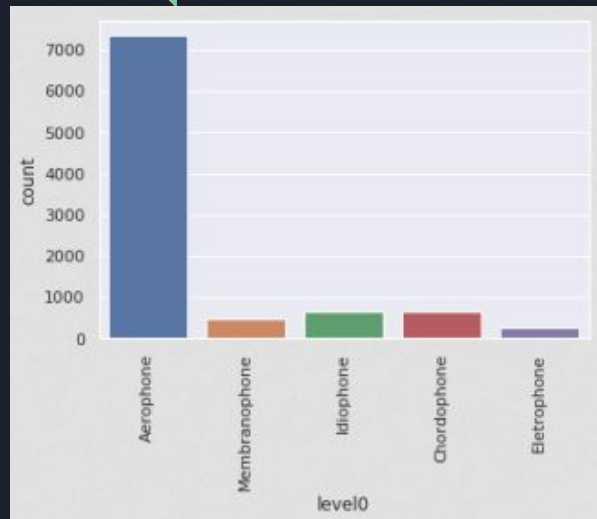
Parmezan, Antonio & Alves de Souza, Vinícius & Batista, Gustavo. (2018).
Towards Hierarchical Classification of Data Streams.



Classificação de Instrumentos

- Não possui nulos
- Extração de 30 features a partir do Mel-frequency cepstrum
 - Características espectrais de tempo curto
 - Baseia-se na escala MEL
 - Apenas o que é perceptível pelo ouvido humano
- Maioria dos dados de aerofones - instrumentos de vento

Classificação de Instrumentos





Ferramentas

- Bibliotecas de classificação flat para a abordagem local - R, scikit-learn, weka, etc
- [Clus](#) -> Implementa o Clus HMC - baseado em Java/Weka - wrapper [python](#)
- [Mulan](#) - Classificação multilabel com apoio para hierárquico - baseado no Weka
- [HDLTex](#): Hierarchical Deep Learning for Text Classification - Python



Para o seminário prático

Avaliar diferentes abordagens:

- Utilizar métricas de avaliação inteligentes
- Avaliar um conjunto de opções dentro da abordagem local
- Verificar a compatibilidade Python de uma abordagem global

Obrigado pela atenção

Bruno Gomes Coelho - 9791160

Pedro Regattieri Rocha - 8531702



[Github](#)