



Classificar documentos é coisa do passado

Henrique Brandão
UAN - Eficiência 1



Motivação

“Quero um modelo de IA para classificar documentos...”

— Alguém, todo dia.



Problemas para resolver

1

Classificar documentos em positivo ou negativo

2

Modelo para poucas classes

3

Interseção entre classes de demandas distintas

4

Baixo reaproveitamento

5

Difícil engajar pessoas na curadoria

6

Não aprendemos com os dados



Objetivo do projeto

Utilizar soluções de **IA** para criar uma **metodologia** de entendimento **ampla** para **documentos**.



Solução proposta

Através de *feature extraction* e *Deep Learning*, podemos relacionar documentos de natureza semelhante e comparar um novo exemplo com os demais sem necessitar de um novo **esforço** de modelagem a cada **demanda**.



Processo

Acesso aos documentos

Já **existem** diversas bases com algum nível de curadoria realizado

Montagem do *dataset*

Consolidar o que foi encontrado de forma **representativa**

Extração de características representativas

Emprego de modelos do tipo Autoencoder para gerar uma melhor **generalização**

Observar os resultados

Modelos de clusterização e redução de **dimensionalidade** podem ser aplicados

Modelo de classificação

Construção do **modelo**, lógica ou heurística para classificar os *data points*

Testar novos documentos

Predições e métricas de **desempenho**



Serviços

- Ganhos **financeiros**
- Ganhos de **eficiência**
- **Pesquisa** e desenvolvimento

Milhões de
documentos

Legado, atuais, novos tipos

Dezenas de
funcis

Qualidade de vida no trabalho

Muitas,
muitas horas

Tarefas mais significativas



Obrigado!



