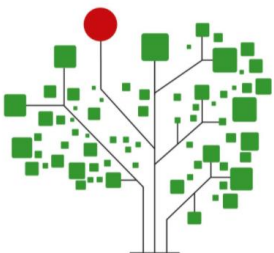




INSTITUTO FEDERAL
Espírito Santo
Campus Serra

Desenvolvimento de Aplicações Inteligentes – IFES/TRE

Projeto de Implementação PLN



Hilário Tomaz Alves de Oliveira

hilario.oliveira@ifes.edu.br

Projeto 1 - Classificação de Decisões Judiciais

Tarefa 1

- Na Aula 6 da disciplina, estudamos sobre a tarefa de **Classificação Automática de Textos**, que é uma das mais comuns da área de **Processamento de Linguagem Natural (PLN)**.
- Implementamos alguns exemplos de **abordagens tradicionais de Aprendizado de Máquina** e **Modelos Neurais de Linguagens** para a tarefa de **classificação automática de decisões judiciais** em documentos escritos em português do Brasil.

Tarefa 1

- Para isso, usamos a **base de dados** disponível em:
 - https://huggingface.co/datasets/joelniklaus/brazilian_court_decisions
- O objetivo deste projeto é **estender os códigos desenvolvidos** em sala para **incorporar novos algoritmos e configurações**.
 1. **Implemente as modificações** indicadas nos códigos.
 2. **Execute os experimentos** com **todos os algoritmos implementados**.
 3. Monte uma **planilha com os resultados gerados**.

Exemplo de Tabela 1

Tabela 1. Resultados do experimentos (%) utilizando a abordagem baseada na medida TF-IDF. O melhor resultado com base em cada medida de avaliação é destacado em negrito.

Algoritmos	Acurácia	Precisão	Cobertura	Medida-F
Árvore de Decisão	75,38	76,18	76,04	76,10
CatBoost	84,13	84,60	84,54	84,53
Extra Trees	76,57	80,18	75,43	77,03
Floresta Aleatória	78,96	82,00	78,21	79,57
KNN	67,82	68,02	70,16	68,86
LightGBM	84,96	85,01	85,82	85,37
MLP	77,73	78,97	78,11	78,50
Regressão Logística	82,14	81,95	83,71	82,69
SVM	82,11	82,03	83,49	82,67
XGBoost	84,80	85,31	85,20	85,22

Exemplo de Tabela 2

Tabela 3. Resultados (%) dos experimentos utilizando os modelos neurais de linguagem pré-treinados. O melhor resultado com base em cada medida de avaliação é destacado em negrito.

Modelos	Épocas	Acurácia	Precisão	Cobertura	Medida-F
ALBERT-base	1	91,823	91,894	91,993	91,900
	5	93,275	93,449	93,322	93,369
	10	92,399	92,658	92,367	92,492
BERT-base	1	92,906	93,117	93,036	93,051
	5	93,552	93,604	93,721	93,645
	10	93,370	93,442	93,401	93,406
DistilBERT-base	1	93,012	93,209	93,109	93,139
	5	93,592	93,737	93,665	93,684
	10	93,490	93,535	93,577	93,543
RoBERTa-base	1	92,695	92,874	92,869	92,842
	5	93,914	93,913	94,035	93,951
	10	93,943	94,181	93,949	94,051

Projeto 2 – Reconhecimento de Entidades Nomeadas

Tarefa 2

- Na Aula 7 da disciplina, estudamos sobre a tarefa de **Reconhecimento de Entidades Nomeadas (REN)**.
 - A ideia dessa tarefa é reconhecer menções de entidades de interesse dentro de um texto.
- Implementamos uma abordagem usando **Redes Neurais Profundas** para a tarefa de **REN** em documentos jurídicos escritos em português do Brasil.

Tarefa 2

- Para isso, usamos a **base de dados** disponível em:
 - https://github.com/messias077/ner_pt/tree/main/data/corpora/le_ner
 - https://huggingface.co/datasets/lener_br
- O objetivo deste projeto é **estender os códigos desenvolvidos** em sala para **incorporar a base de dados ulysses-ner-br**.
 - https://github.com/ulysses-camara/ulysses-ner-br/tree/main/annotated-corpora/PL_corpus_conll/pl_corpus_categorias

Tarefa 2

- Incorpore a base de dados anterior no código.
- Altere os parâmetros a seguir para:
 - `max_iterations=100`
 - `all_possible_transitions=True`
- Execute os experimentos para as duas bases de dados (LE-NER e ulysses-ner-br).
- Por fim, monte uma planilha com os resultados gerados.

Exemplo de Tabela 2

Entidade	ES			MG			RJ			SP		
	P	C	F1	P	C	F1	P	C	F1	P	C	F1
ARTIGO	98,43	98,43	98,43	97,38	97,78	97,58	94,29	97,70	95,96	96,14	97,33	96,73
DECRETO	25,00	27,27	26,09	100,00	100,00	100,00	96,00	97,56	96,77	91,89	96,23	94,01
INCISO	87,18	94,44	90,67	80,65	97,40	88,24	84,38	96,43	90,00	95,74	98,54	97,12
LEI	99,45	99,45	99,45	95,01	98,07	96,52	98,57	98,92	98,74	97,32	96,24	96,77
LOCAL	70,45	43,06	53,45	74,63	40,65	52,63	55,47	53,94	54,69	32,00	29,88	30,90
ORGANIZAÇÃO	67,99	59,47	63,44	66,67	56,78	61,33	55,49	40,31	46,70	58,52	58,61	58,57
PORTARIA	66,67	66,67	66,67	19,64	91,67	32,35	95,24	71,43	81,63	72,22	81,25	76,47
TEMPO	39,42	53,25	45,30	82,09	90,94	86,29	87,83	95,83	91,66	84,55	89,42	86,92

<https://repositorio.ifes.edu.br/handle/123456789/2944>

Entrega dos Resultados e Códigos

- Cada grupo deve realizar suas implementações e gerar seus resultados para as duas tarefas (Classificação e REN).
 - Gere um tabela para cada Tarefa.
- Compartilhar a planilha com resultados e os códigos em um repositório do Github.
 - O link deve ser enviado por e-mail para: hilariotomaz@gmail.com
 - **OBS:** No e-mail devem contar os nomes dos membros da equipe.
 - Data Entrega: 13/12/2023