



# FORMAÇÃO INTELIGÊNCIA ARTIFICIAL E MACHINE LEARNING

---

NLP

Prof. Fernando Amaral – Todos os Direitos Reservados

# UD

---

- Projeto Open Souce
- Banco de dados para mais de 60 idiomas
- <http://universaldependencies.org/>



# CoNLL-U

Formato de padrão para annotations

<http://universaldependencies.org/format.html>

doc_id	Identificador do documento (corpus com vários documentos)	Inteiro
paragraph_id	identifica o paragrafo dentro do documento	Inteiro
sentence_id	Identificador da Sentença	Inteiro
sentence	Sentença	
token_id	Identificador do Token	Inteiro
token	Token	
lemma	Lemmatization	
upos	Tipo de palavra	
xpos	Tipo de palavra, mas específico do idioma	
feats	Características morfológicas do idioma	
head_token_id	Cabeçalho da palavra	
dep_rel	Relação de dependência	
deps	Relação de dependência da forma de pares	
misc	Aspectos diversos	



# Pacote UDpipe

---

- Biblioteca para C++, Java, C#, Python, R
- Tokenization, POS Tagging, Lemmatization, Dependency Parsing
- Sem dependências externas (ex. Java)
- Funções de anotação no formato **CoNLL-U Format**
  - <http://universaldependencies.org/format.html#conll-u-format>
- Modelo em português ( e outros 50 idiomas)
- Você pode criar seu próprio modelo



Machine learning is a field of computer science that gives computer systems the ability to learn (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed.

Editor de dados

Arquivo Editar Ajuda

	doc_id	paragraph_id	sentence_id	sentence	token_id	token	lemma	upos	xpos	feats	head_token_id	dep_rel	deps	misc
1	1	1	1	Machine learning is a field of computer science t>	1	Machine	machine	NOUN	NN	Number=Sing	2	compound		
2	1	1	1	Machine learning is a field of computer science t>	2	learning	learn	NOUN	NN	Number=Sing	5	nsubj		
3	1	1	1	Machine learning is a field of computer science t>	3	is	be	AUX	VBZ	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=	5	cop		
4	1	1	1	Machine learning is a field of computer science t>	4	a	a	DET	DT	Definite=Ind PronType=Art	5	det		
5	1	1	1	Machine learning is a field of computer science t>	5	field	field	NOUN	NN	Number=Sing	0	root		
6	1	1	1	Machine learning is a field of computer science t>	6	of	of	ADP	IN		8	case		
7	1	1	1	Machine learning is a field of computer science t>	7	computer	computer	NOUN	NN	Number=Sing	8	compound		
8	1	1	1	Machine learning is a field of computer science t>	8	science	science	NOUN	NN	Number=Sing	5	nmod		
9	1	1	1	Machine learning is a field of computer science t>	9	that	that	PRON	WDT	PronType=Rel	10	nsubj		
10	1	1	1	Machine learning is a field of computer science t>	10	gives	give	VERB	VBZ	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=	5	acl:relcl		
11	1	1	1	Machine learning is a field of computer science t>	11	computer	computer	NOUN	NN	Number=Sing	12	nsubj		
12	1	1	1	Machine learning is a field of computer science t>	12	systems	system	VERB	VBZ	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=	10	ccomp		
13	1	1	1	Machine learning is a field of computer science t>	13	the	the	DET	DT	Definite=Def PronType=Art	14	det		
14	1	1	1	Machine learning is a field of computer science t>	14	ability	ability	NOUN	NN	Number=Sing	12	obj		
15	1	1	1	Machine learning is a field of computer science t>	15	to	to	PART	TO		16	mark		
16	1	1	1	Machine learning is a field of computer science t>	16	learn	learn	VERB	VB	VerbForm=Inf	14	acl		
17	1	1	1	Machine learning is a field of computer science t>	17	(	(	PUNCT	-LRB-		21	punct		SpaceAfter=No
18	1	1	1	Machine learning is a field of computer science t>	18	i.e.	i.e.	X	FW		21	advmod		SpaceAfter=No
19	1	1	1	Machine learning is a field of computer science t>	19	,	,	PUNCT	,		21	punct		
20	1	1	1	Machine learning is a field of computer science t>	20	progressively	progressively	ADV	RB		21	advmod		
21	1	1	1	Machine learning is a field of computer science t>	21	improve	improve	VERB	VB	VerbForm=Inf	16	parataxis		
22	1	1	1	Machine learning is a field of computer science t>	22	performance	performance	NOUN	NN	Number=Sing	21	obj		
23	1	1	1	Machine learning is a field of computer science t>	23	on	on	ADP	IN		26	case		
24	1	1	1	Machine learning is a field of computer science t>	24	a	a	DET	DT	Definite=Ind PronType=Art	26	det		
25	1	1	1	Machine learning is a field of computer science t>	25	specific	specific	ADJ	JJ	Degree=Pos	26	amod		
26	1	1	1	Machine learning is a field of computer science t>	26	task	task	NOUN	NN	Number=Sing	21	obl		SpaceAfter=No
27	1	1	1	Machine learning is a field of computer science t>	27	)	)	PUNCT	-RRB-		21	punct		
28	1	1	1	Machine learning is a field of computer science t>	28	with	with	ADP	IN		29	case		
29	1	1	1	Machine learning is a field of computer science t>	29	data	data	NOUN	NN	Number=Sing	16	obl		SpaceAfter=No
30	1	1	1	Machine learning is a field of computer science t>	30	,	,	PUNCT	,		5	punct		
31	1	1	1	Machine learning is a field of computer science t>	31	without	without	SCONJ	IN		34	mark		
32	1	1	1	Machine learning is a field of computer science t>	32	being	be	AUX	VBG	VerbForm=Ger	34	aux:pass		
33	1	1	1	Machine learning is a field of computer science t>	33	explicitly	explicitly	ADV	RB		34	advmod		
34	1	1	1	Machine learning is a field of computer science t>	34	programmed	program	VERB	VCN	Tense=Past VerbForm=Part	5	advcl		SpaceAfter=No
35	1	1	1	Machine learning is a field of computer science t>	35	.	.	PUNCT	.		5	punct		SpacesAfter=\\n

