TRABALHO 1 - AM2

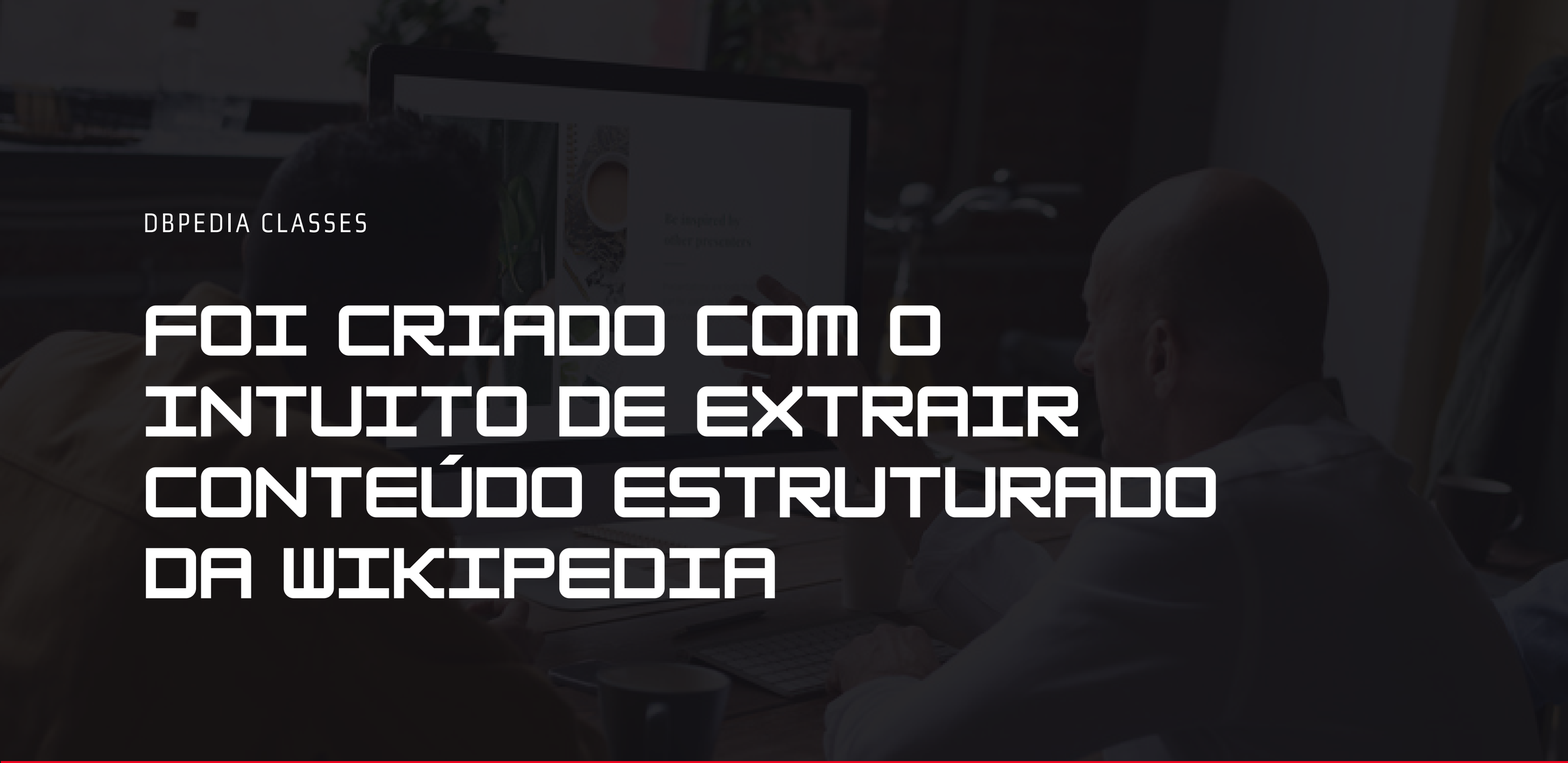
B D PEDIA CLASSES

ANA ELLEN DEODATO P SILVA 800206

VINICIUS DE OLIVEIRA GUIMARÃES 802431

VINÍCIUS GONÇALVES PERILLO 800219

SUMÁRIO



DBPEDIA CLASSES

FOI CRIADO COM O INTUITO DE EXTRAIR CONTEÚDO ESTRUTURADO DA WIKIPEDIA

ANÁLISE DOS DADOS

```
[25] data_train.sample(10)
```

	text	11	12	13
109080	1909 Alekhin, provisional designation 1972 RW2...	Place	CelestialBody	Planet
170932	The 1989 U.S. Women's Open Golf Championship w...	Event	Tournament	GolfTournament
43991	Cameroon Airlines was an airline from Cameroon...	Agent	Company	Airline
131746	Jam Mir Mohammad Yousaf Aliani (Urdu: جام میر ...	Agent	Politician	President
[illegible]	H.O.W. Journal is a bi-annual non-profit art &...	Work	PeriodicalLiterature	Magazine
17815	Mustapha Khaznadar (1817–1878 ,مصطفی خزندار), ...	Agent	Politician	PrimeMinister
49880	Xenia Knoll (born 2 September 1992 in Biel) is...	Agent	Athlete	TennisPlayer
88971	Buvik Church (Norwegian: Buvik kirke) is a par...	Place	Building	HistoricBuilding
106575	Buen Retiro Ballivian Airport (ICAO: SLBT) is ...	Place	Infrastructure	Airport
87557	Llyn Melynllyn (Welsh for yellow lake) is a l...	Place	BodyOfWater	Lake

TEXTOS DA WIKIPEDIA

CATEGORIAS POR NÍVEL

PRECISA DE PLN!!

PRÉ-PORCESSAMENTO

PLN NOS DADOS!

PRÉ-PROCESSAMENTO

PNL NOS DADOS!

1 TOKENIZADOS

2 REMOVIDAS AS STOPWORDS

3 LEMMATIZADOS

4 WORD EMBENDING

PRÉ-PROCESSAMENTO

PNL NOS DADOS

text	l1	l2	l3	sep	tokens
William Alexander Massey (October 7, 1856 – Ma...	Agent	Politician	Senator	train	[William, Alexander, Massey, (, October, 7, ,,...
Lions is the sixth studio album by American ro...	Work	MusicalWork	Album	train	[Lions, sixth, studio, album, American, rock, ...
Pirqa (Aymara and Quechua for wall, hispaniciz...	Place	NaturalPlace	Mountain	train	[Pirqa, (, Aymara, Quechua, wall, ,, hispanici...
Cancer Prevention Research is a biweekly peer-...	Work	PeriodicalLiterature	AcademicJournal	train	[Cancer, Prevention, Research, biweekly, peer-...
The Princeton University Chapel is located on ...	Place	Building	HistoricBuilding	train	[Princeton, University, Chapel, located, unive...

```
array([-0.0601205 ,  0.203194  , -0.22487924, -0.20244634,  0.11716624,
        0.297426  , -0.29432997, -0.4024878 , -1.0120828 , -0.3733864 ,
        0.3730548 , -0.01706478, -0.24110077, -0.2525653 ,  0.60188925,
       -0.4820766 , -0.3772207 ,  0.13282822, -1.04063   ,  0.47232217,
        0.0583109 , -0.38869864,  0.37759513,  0.25988442, -0.19599321,
       -0.9853721 ,  0.09595042, -0.59779793, -0.27062252,  0.3343564 ,
        2.5358407 , -0.39028415, -0.5450571 , -0.05943721, -0.03535265,
       -0.15817885,  0.38572928,  0.30169192,  0.2726746 , -0.02003207,
        0.19918783, -0.02332716,  0.16651492, -0.62228787, -0.4012342 ,
        0.44148806, -0.19441378, -0.7447943 ,  0.29198354,  0.07799093]
      dtype=float32),
```

NOSSOS MÉTODOS

CLASSIFICAÇÃO

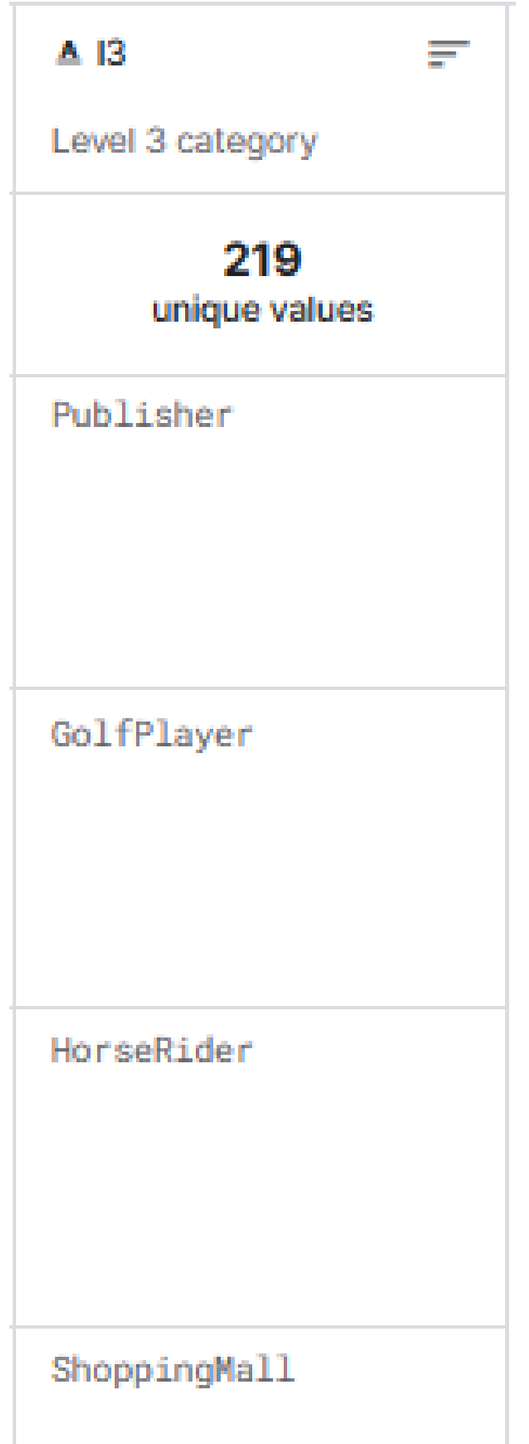
PLANA

CLASSIFICAÇÃO

LOCAL POR PAI

CLASSIFICAÇÃO PLANA

1. VAMOS CONSIDERAR APENAS A COLUNA L3
2. A PARTIR DA CLASSE L3, CONSEGUIMOS ENCONTRAR AS SUPERIORES
3. REGRESSÃO LOGÍSTICA X RANDOM FOREST



MÉTRICAS

PLANA - REGRESSÃO LOGÍSTICA

50 DIMENSÕES

PRECISION : 0.681358357732671

RECALL: 0.908477810310228

F-SCORE: 0.7786952659801953

MÉTRICAS

PLANA - RANDOM FOREST

50 DIMENSÕES

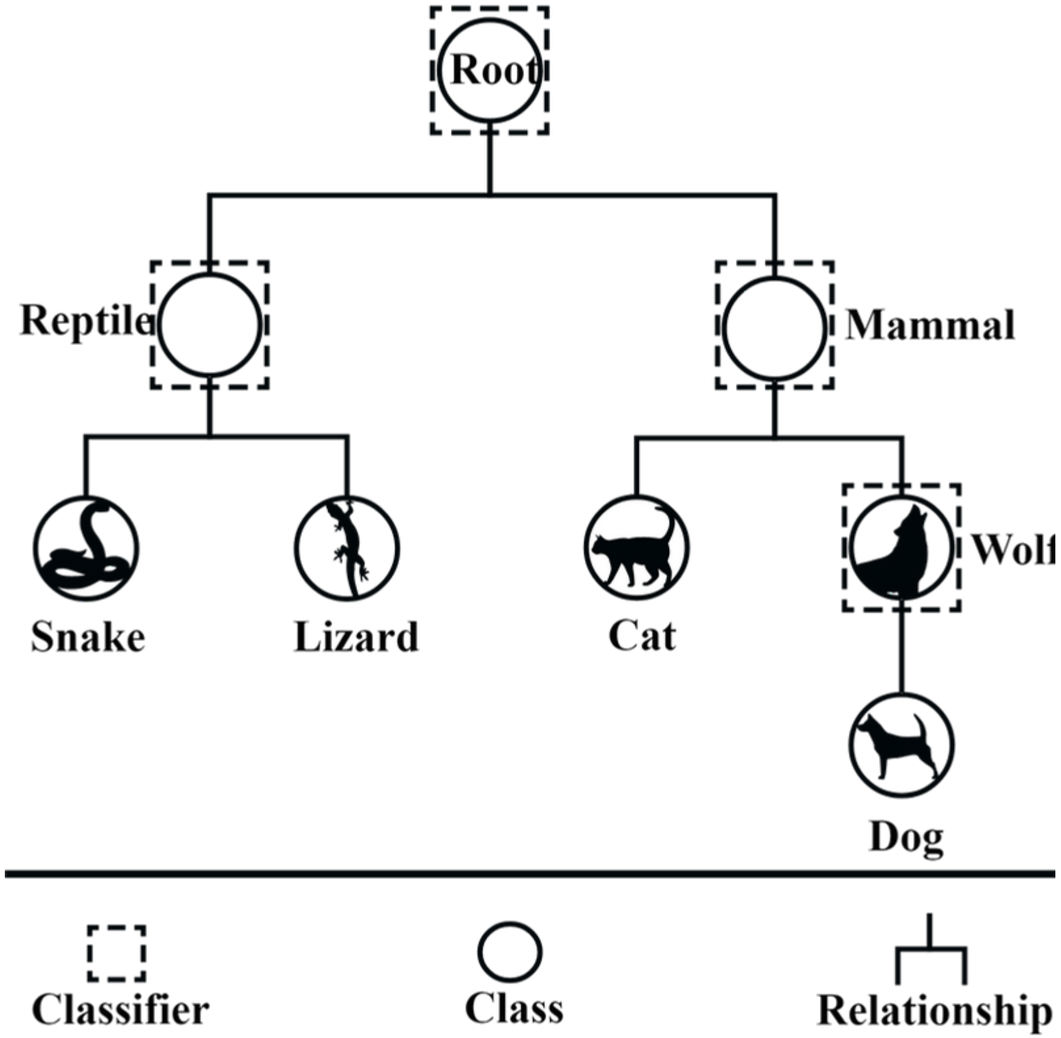
F-SCORE PL50: 0.8848954392429077

PRECISION PL50: 0.8848954392429077

RECALL PF50: 0.8364421927602504

CLASSIFICAÇÃO LOCAL POR PAI

1. LOCAL CLASSIFIER PARENT NODE - HICLASS
2. TREINO EM PARALELO PREDIÇÃO TOP-DOWN
3. REGRESSÃO LOGÍSTICA X RANDOM FOREST



CLASSIFICAÇÃO
LOCAL POR NÓ PAI

	0	0	1	1	2	2
0	Agent	Agent	Person	Politician	OfficeHolder	PrimeMinister
1	Agent	Agent	Actor	Actor	AdultActor	AdultActor
2	Place	Place	Infrastructure	Infrastructure	Dam	Dam
3	Event	Event	Race	Race	CyclingRace	CyclingRace
4	Agent	Agent	Athlete	Athlete	MartialArtist	MartialArtist
...	...	...	...	...	...	...
60789	Agent	Agent	Person	Person	Economist	BusinessPerson
60790	Agent	Agent	SportsTeam	SportsLeague	RugbyClub	SoccerLeague
60791	Agent	Agent	Athlete	Athlete	GolfPlayer	GolfPlayer
60792	Agent	Agent	Athlete	Athlete	AustralianRulesFootballPlayer	AustralianRulesFootballPlayer
60793	Place	Place	Building	Building	Castle	Castle

MÉTRICAS

LOCAL POR PAI - REGRESSÃO LOGÍSTICA

300 DIMENSÕES

F-SCORE PL300: 0.7786952659801953

PRECISION PL300: 0.681358357732671

RECALL PL300: 0.908477810310228

50 DIMENSÕES

F-SCORE PL50: 0.7959063942713644

PRECISION PL50: 0.7959063942713644

RECALL PL50: 0.7959063942713644

MÉTRICAS

LOCAL POR PAI - RANDOM FOREST

300 DIMENSÕES

F-SCORE PF300: 0.7445189609876915

PRECISION PF300: 0.65145409086423

RECALL PF300: 0.8686054544856401

50 DIMENSÕES

F-SCORE PF50: 0.8398252020484478

PRECISION PF50: 0.8398252020484478

RECALL PF50: 0.8398252020484478

CONCLUSÃO

COMPARANDO AS MÉTRICAS ...

CONHEÇA NOSSA EQUIPE

ANA ELLEN

EREN

VINI PERILLO

PELINHO

VINI GUIMARÃES

GUIMA


OBRIGADO!