A decorative graphic on the left side of the slide, consisting of a grid of squares in various shades of green and one red square. The squares are arranged in a non-uniform pattern, with some being larger than others. The colors range from dark forest green to light lime green, with one bright red square in the middle-left area.

# Reconhecimento de Emoções em Tweets

Kevila Morais

# Objetivo

- Realizar uma análise comparativa entre diferentes abordagens de aprendizado de máquina para identificação de emoções.
- MACHINE LEARNING
  - Logistic Regression
  - Multinomial NB
  - KNN
  - Decision tree
  - Random Forest
- DEEP LEARNING
  - Modelo com Embedding
  - CNN
  - LSTM

# Repositório Corpus

sadness (0), joy (1), love (2), anger (3), fear (4), surprise (5)

text (string)	label (class label)
"i didnt feel humiliated"	0 (sadness)
"i can go from feeling so hopeless to so damned hopeful just from being around someone who cares and is awake"	0 (sadness)
"im grabbing a minute to post i feel greedy wrong"	3 (anger)
"i am ever feeling nostalgic about the fireplace i will know that it is still on the property"	2 (love)
"i am feeling grouchy"	3 (anger)
"ive been feeling a little burdened lately wasnt sure why that was"	0 (sadness)
"ive been taking or milligrams or times recommended amount and ive fallen asleep a lot faster but i also feel like so funny"	5 (surprise)
"i feel as confused about life as a teenager or as jaded as a year old man"	4 (fear)
"i have been with petronas for years i feel that petronas has performed well and made a huge profit"	1 (joy)

<https://huggingface.co/datasets/dair-ai/emotion>

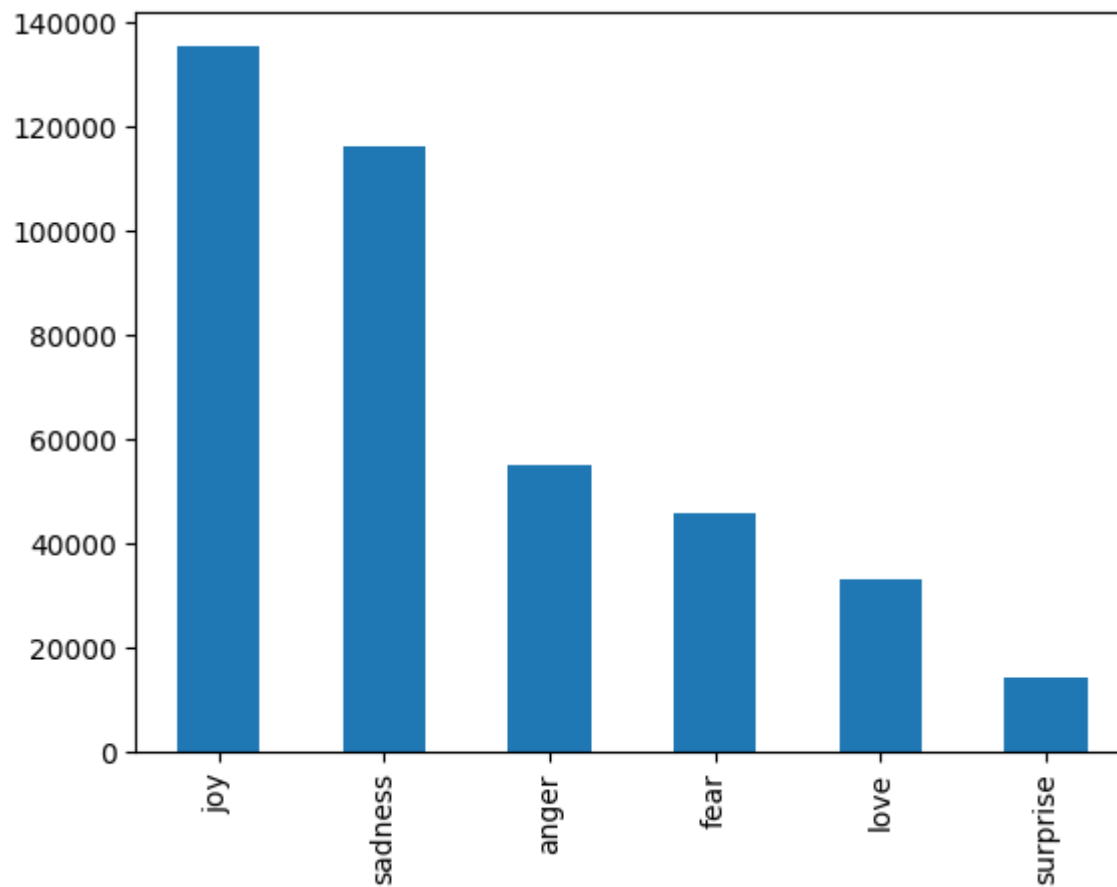
# Repositório Corpus

name	train	validation	test
split	16000	2000	2000
unsplit	416809	n/a	n/a

<https://huggingface.co/datasets/dair-ai/emotion>

# Repositório Corpus

Quantidade de textos por rótulo



# Trabalhos Relacionados

# Processamento do texto

- en\_core\_news\_sm
- Tradução de emoji
- Lematização
- Normalização de acentos
- Substituição de caracteres especiais

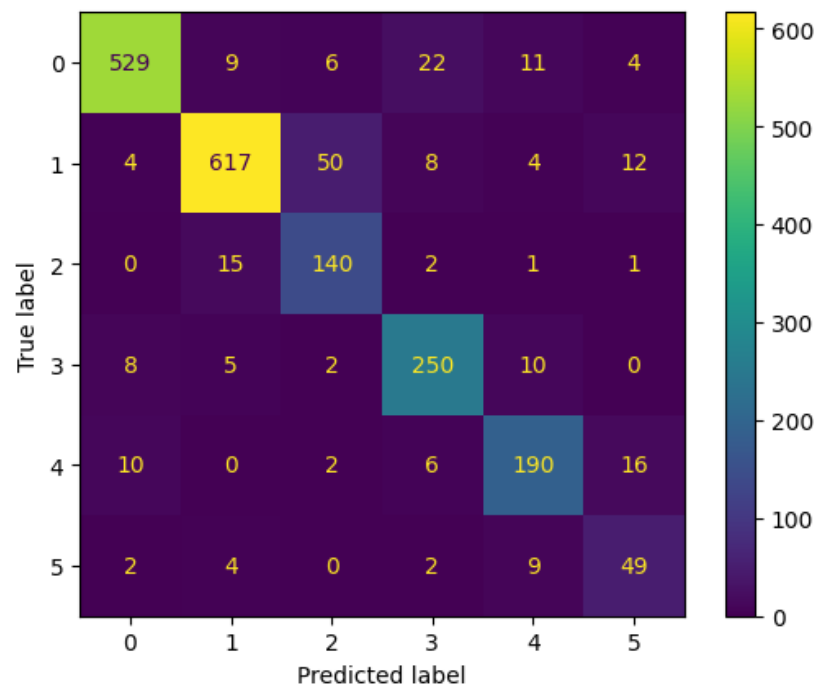
# Vetorização

- `CountVectorizer(binary=True, max_features=None, ngram_range=(1, 1))`
- Vocabulário: 15700 palavras



# Resultados treinamentos Machine Learning

## Logistic Regression

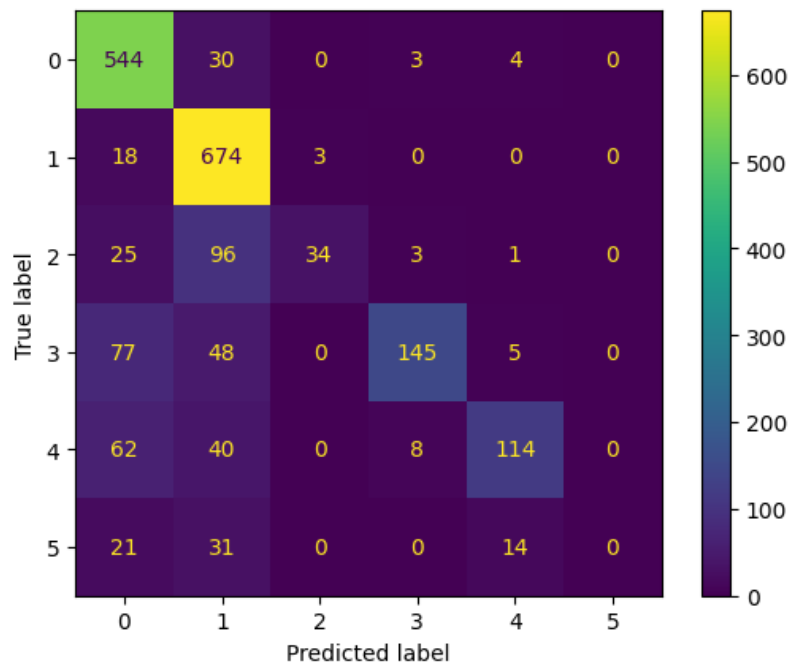


Classifier: Logistic Regression

	precision	recall	f1-score	support
0	0.96	0.91	0.93	581
1	0.95	0.89	0.92	695
2	0.70	0.88	0.78	159
3	0.86	0.91	0.88	275
4	0.84	0.85	0.85	224
5	0.60	0.74	0.66	66
accuracy			0.89	2000
macro avg	0.82	0.86	0.84	2000
weighted avg	0.90	0.89	0.89	2000

# Resultados treinamentos Machine Learning

## Multinomial NB

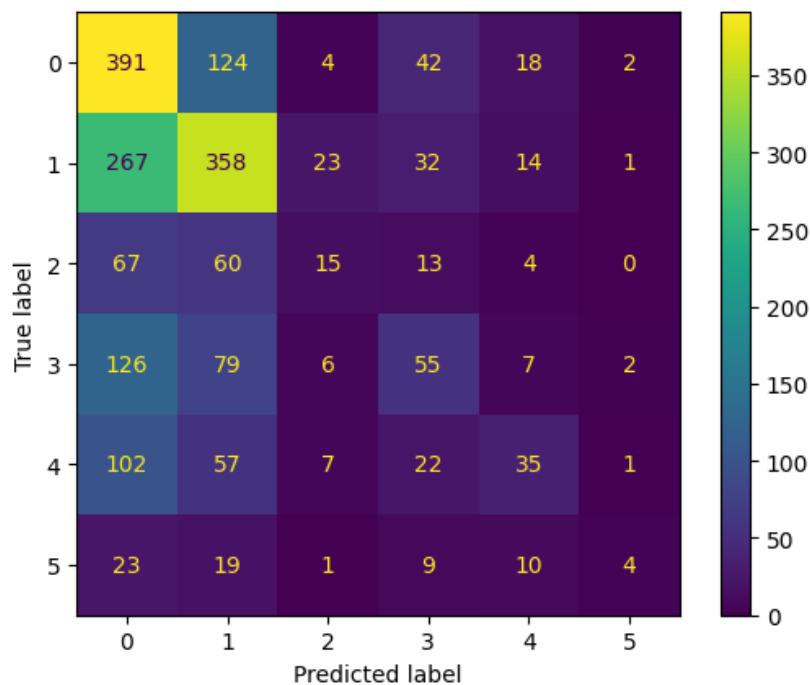


Classifier: Multinomial NB

	precision	recall	f1-score	support
0	0.73	0.94	0.82	581
1	0.73	0.97	0.84	695
2	0.92	0.21	0.35	159
3	0.91	0.53	0.67	275
4	0.83	0.51	0.63	224
5	0.00	0.00	0.00	66
accuracy			0.76	2000
macro avg	0.69	0.53	0.55	2000
weighted avg	0.76	0.76	0.72	2000

# Resultados treinamentos Machine Learning

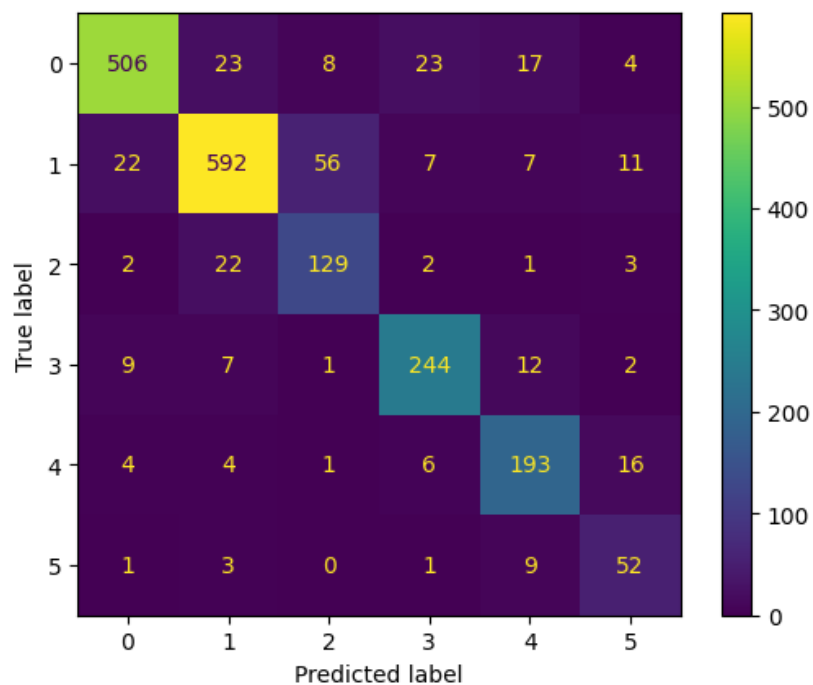
## KNN



	precision	recall	f1-score	support
0	0.40	0.67	0.50	581
1	0.51	0.52	0.51	695
2	0.27	0.09	0.14	159
3	0.32	0.20	0.25	275
4	0.40	0.16	0.22	224
5	0.40	0.06	0.11	66
accuracy			0.43	2000
macro avg	0.38	0.28	0.29	2000
weighted avg	0.42	0.43	0.40	2000

# Resultados treinamentos Machine Learning

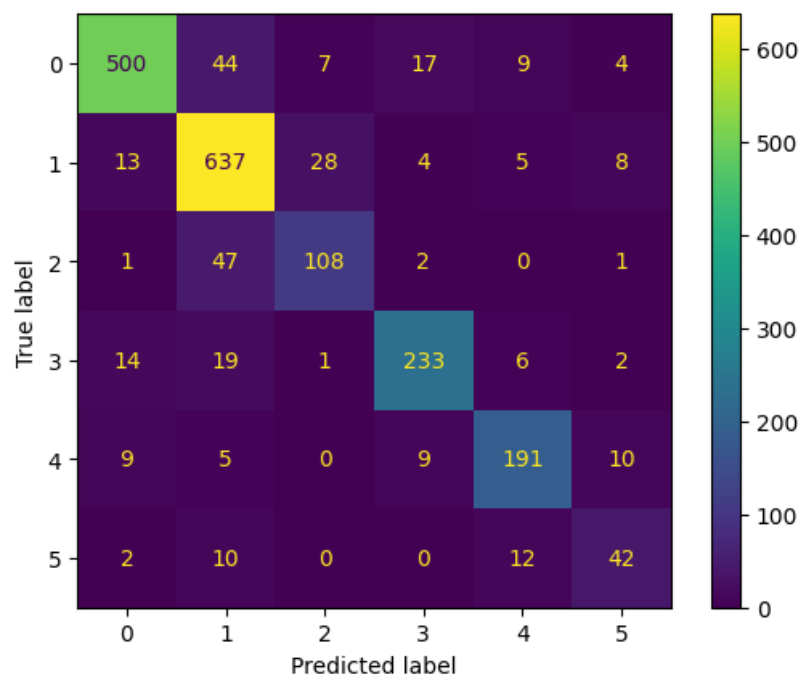
## Decision Tree



	precision	recall	f1-score	support
0	0.93	0.87	0.90	581
1	0.91	0.85	0.88	695
2	0.66	0.81	0.73	159
3	0.86	0.89	0.87	275
4	0.81	0.86	0.83	224
5	0.59	0.79	0.68	66
accuracy			0.86	2000
macro avg	0.79	0.85	0.82	2000
weighted avg	0.87	0.86	0.86	2000

# Resultados treinamentos Machine Learning

## Random Forest



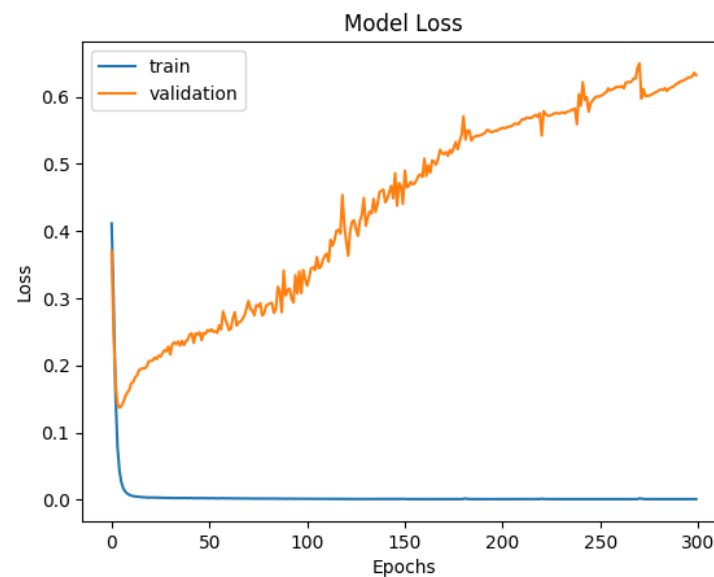
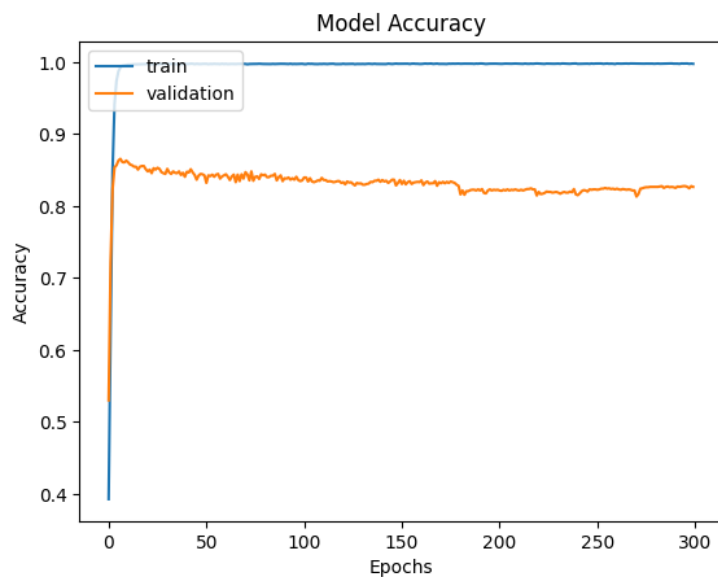
	precision	recall	f1-score	support
0	0.93	0.86	0.89	581
1	0.84	0.92	0.87	695
2	0.75	0.68	0.71	159
3	0.88	0.85	0.86	275
4	0.86	0.85	0.85	224
5	0.63	0.64	0.63	66
accuracy			0.86	2000
macro avg	0.81	0.80	0.80	2000
weighted avg	0.86	0.86	0.86	2000

# Resultados treinamentos Deep Learning

## Modelo simples com Embedding

1,948,654 parâmetros

300 épocas

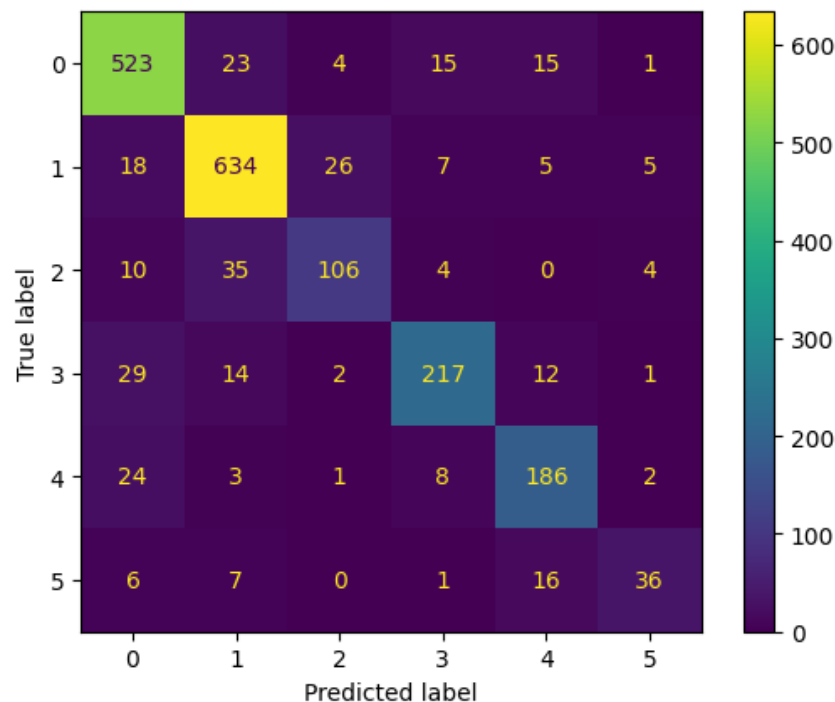


# Resultados treinamentos Deep Learning

## Modelo simples com Embedding

1,948,654 parâmetros

300 epochs



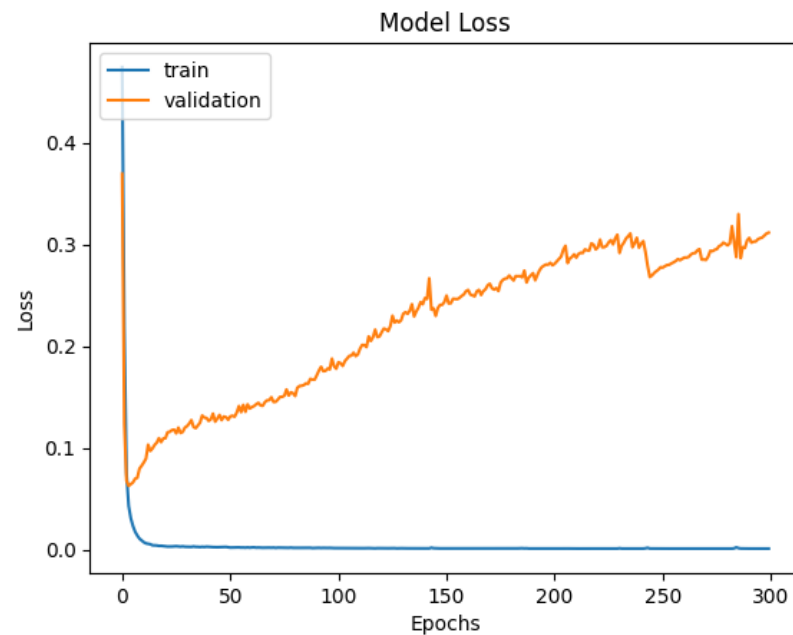
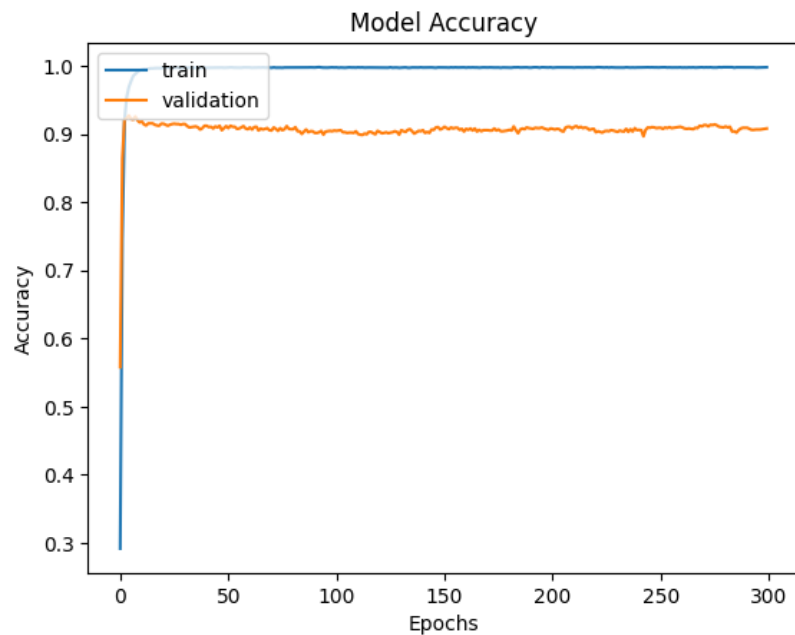
	precision	recall	f1-score	support
0	0.86	0.90	0.88	581
1	0.89	0.91	0.90	695
2	0.76	0.67	0.71	159
3	0.86	0.79	0.82	275
4	0.79	0.83	0.81	224
5	0.73	0.55	0.63	66
accuracy			0.85	2000
macro avg	0.82	0.77	0.79	2000
weighted avg	0.85	0.85	0.85	2000

# Resultados treinamentos Deep Learning

## CNN

1,582,734 parâmetros

300 épocas



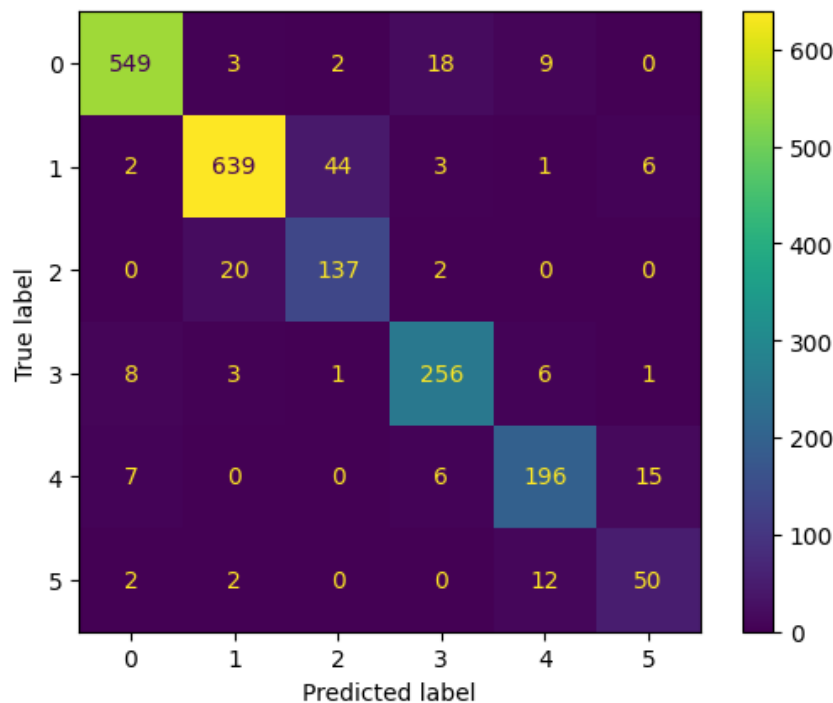


# Resultados treinamentos Deep Learning

## CNN

1,582,734 parâmetros

300 épocas



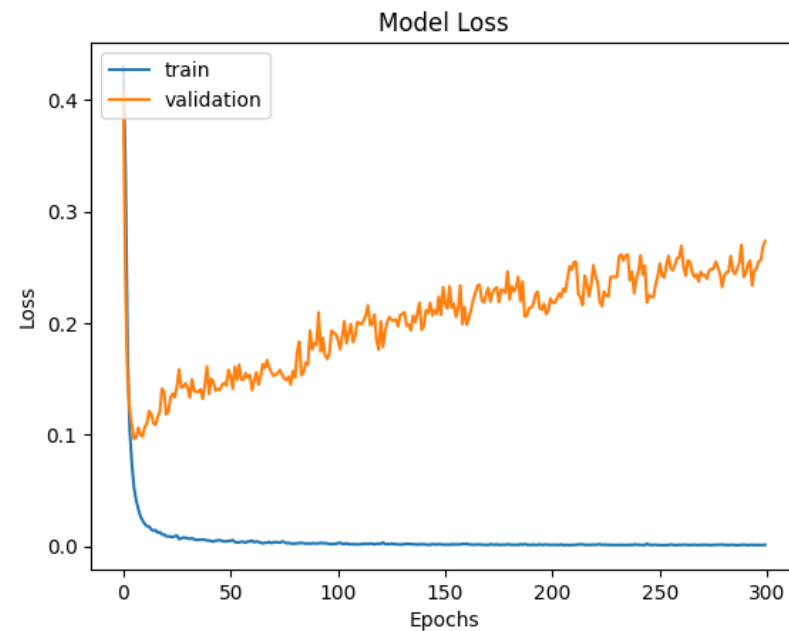
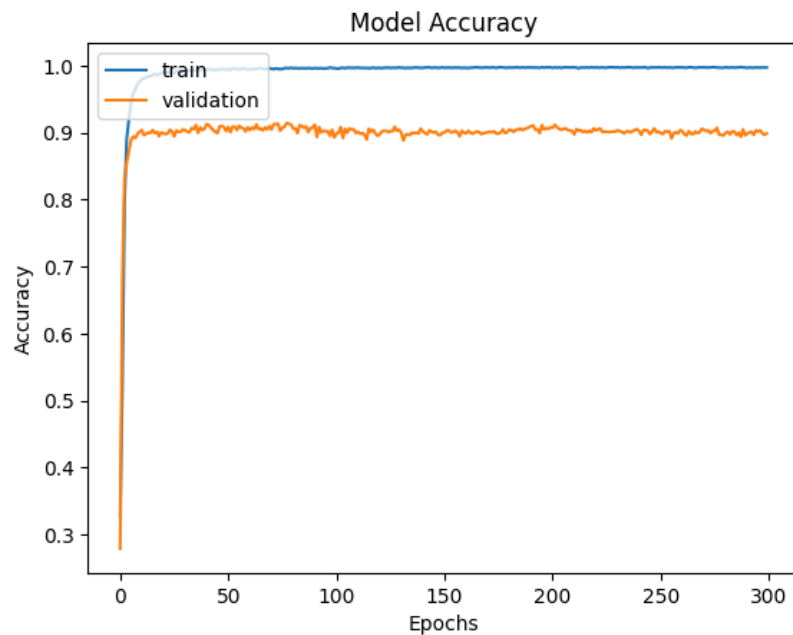
	precision	recall	f1-score	support
0	0.97	0.94	0.96	581
1	0.96	0.92	0.94	695
2	0.74	0.86	0.80	159
3	0.90	0.93	0.91	275
4	0.88	0.88	0.88	224
5	0.69	0.76	0.72	66
accuracy			0.91	2000
macro avg	0.86	0.88	0.87	2000
weighted avg	0.92	0.91	0.91	2000

# Resultados treinamentos Deep Learning

## LSTM

1,821,934 parâmetros

300 épocas

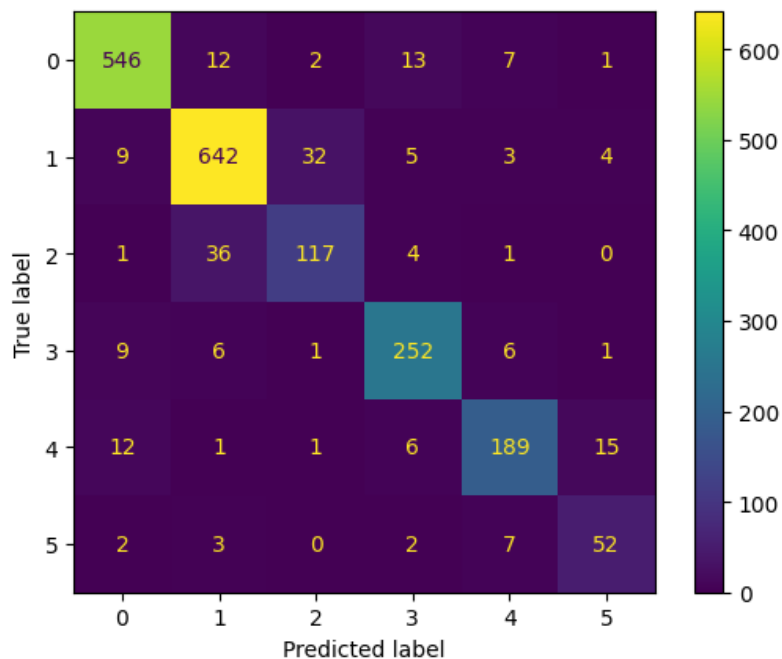


# Resultados treinamentos Deep Learning

## LSTM

1,821,934 parâmetros

300 épocas



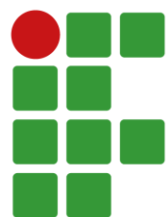
	precision	recall	f1-score	support
0	0.94	0.94	0.94	581
1	0.92	0.92	0.92	695
2	0.76	0.74	0.75	159
3	0.89	0.92	0.90	275
4	0.89	0.84	0.86	224
5	0.71	0.79	0.75	66
accuracy			0.90	2000
macro avg	0.85	0.86	0.85	2000
weighted avg	0.90	0.90	0.90	2000

# Comparação entre acurácia do teste de cada técnica

	TÉCNICA	ACURÁCIA
MACHINE LEARNING	Logistic Regression	<b>0,89</b>
	Multinomial NB	0,76
	KNN	0,43
	Decision Tree	0,86
	Random Forest	0,86
DEEP LEARNING	Modelo com Embedding	0,85
	CNN	<b>0,91</b>
	LSTM	0,90

# Comparação entre resultados das técnicas por classe

	Logistic Regression	Multinomial NB	KNN	Decision Tree	Random Forest	Modelo com Embedding	CNN	LSTM
sadness (0)	529	544	391	506	500	532	549	546
joy (1)	617	674	358	592	637	634	639	642
love (2)	140	34	15	129	108	106	137	117
anger (3)	250	145	55	244	233	217	256	252
fear (4)	190	114	35	193	191	186	196	189
surprise (5)	49	0	4	52	42	36	50	52



**INSTITUTO FEDERAL**

Espírito Santo

Educação pública, gratuita e de qualidade