



Programa de Pós-graduação em Informática - PPGI
Departamento de Ciências da Computação
Universidade de Brasília

BRAVO: Brazilian Portuguese Emotional Intensity Recognition Assistant for Voice

Aluno: Henrique Tibério Brandão Vieira Augusto (221101092)

Orientador: Geraldo Pereira de Rocha Filho

27 de Abril de 2023

Sumário



- 1) Contextualização
- 2) Problema de pesquisa
- 3) Objetivo
- 4) Trabalhos Relacionados
- 5) BRAVO
- 6) Cronograma
- 7) Resultados preliminares



1) Contextualização

Contextualização



Contextualização



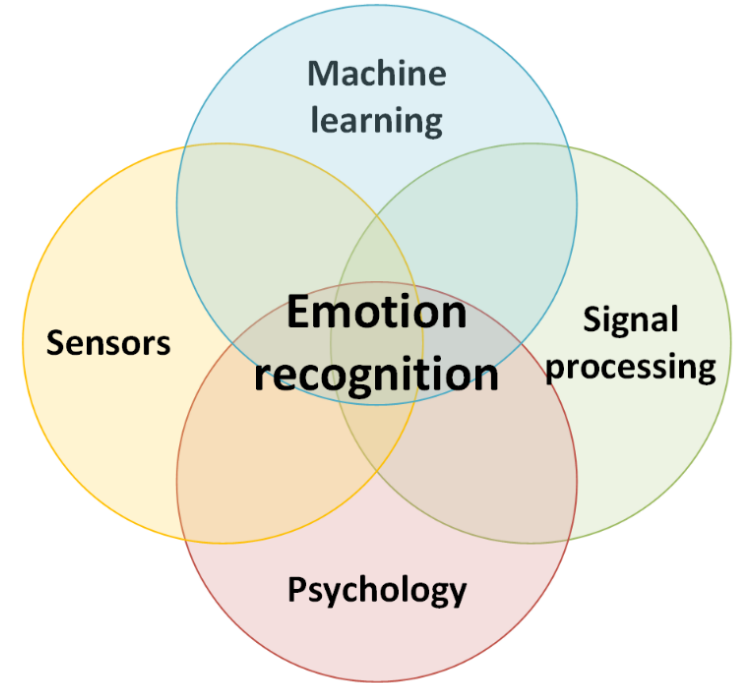
- Reconhecimento de Emoção na voz (*Speech Emotion Recognition, SER*) é um problema **complexo**, pois a **expressão emocional** depende de diversas características (e.g.: linguagem falada, dialeto, sotaque).



Contextualização



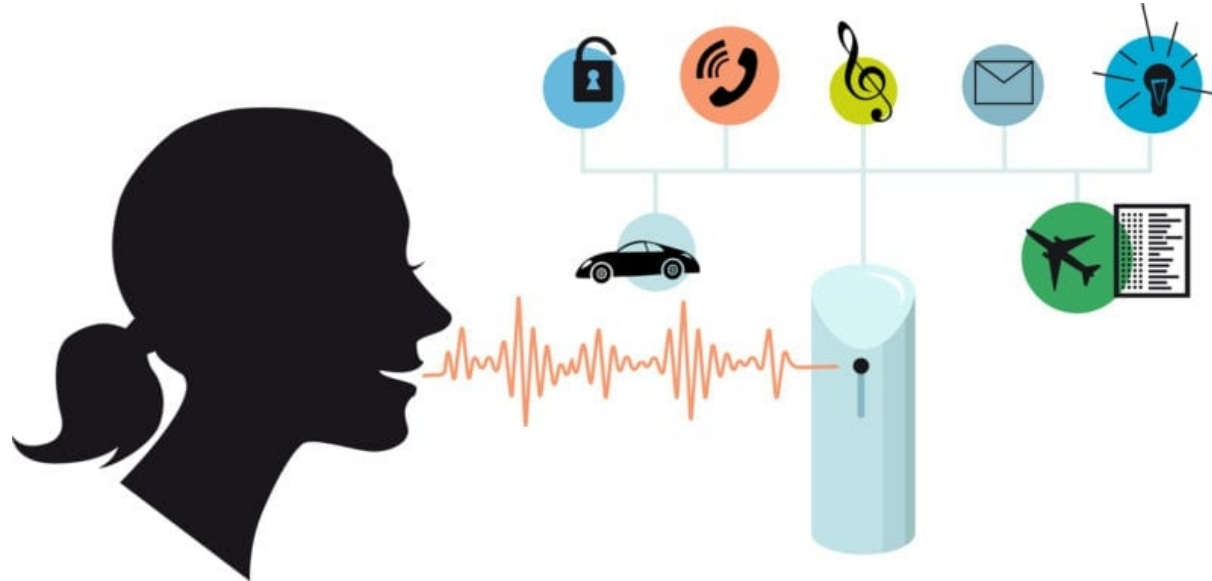
- O Aprendizado de Máquina é uma subárea da **Inteligência Artificial** que visa criar um sistema que receba **informações** para realizar uma tarefa.
- SER se tornou um área de pesquisa proeminente, muito graças ao aumento da **capacidade computacional** e a eficiência de algoritmos.



Contextualização



- Inferir a intensidade de uma emoção tem diversas áreas de **aplicação** para interações humano-computador: **saúde**, **segurança**, **comercial** e de **entretenimento**.





2) Problema de pesquisa

Problema de pesquisa



- Pesquisando o estado da arte da literatura relacionada, encontramos diversos trabalhos lidando com SER, embora **poucos** trabalhos abordando a **intensidade** dessas emoções.
- Realizando SER em nosso **idioma nativo**, também encontramos poucos trabalhos, embora nenhum destes tratando da **intensidade** das emoções.
- Portanto, haveria a possibilidade deste trabalho **colaborar** para o estado da arte de SER em Português, uma vez que ataca diretamente a inferência da intensidade das emoções em nosso idioma nativo.



3) Objetivo

Objetivo



- Criar uma arquitetura de aprendizado de máquina para classificar a intensidade de emoções no idioma Português Brasileiro. Até então não encontrada na literatura.



4) Trabalhos relacionados

Trabalhos relacionados



- No âmbito das Ciências da Computação, o processamento e reconhecimento de voz é uma área de pesquisa **ativa**, com publicações datando desde o final do século XX (1991 e 1995).

Trabalhos relacionados



- Ensemble methods for spoken emotion recognition in call-centers (2007) [65]:
 - **Ubiquidade** de sistemas automatizados e necessidade de aprimorar a naturalidade das interações humano-computador;
 - Importância de interpretar com **precisão** informações emocionais, podendo colaborar com o atendimento de clientes em chamadas de acordo com a **urgência** percebida.

Trabalhos relacionados



- Improving Automatic Emotion Recognition from speech using Rhythm and Temporal feature (2013) [14]:
 - Relaciona a melhora no desempenho dos modelos com a adição ou combinação de mais *features*. Porém, observou os melhores resultados de quando utilizou apenas **MFCCs**.

Trabalhos relacionados



- Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching (2018) [62]:
 - Utiliza rede CNN para efetuar a extração de características a partir de features espectrais;
 - Utiliza mais de um *dataset* em sua tarefa, aponta a quantidade limitada de **amostras** disponíveis para treinamento.

Trabalhos relacionados



- UNSUPERVISED LEARNING APPROACH TO FEATURE ANALYSIS FOR AUTOMATIC SPEECH EMOTION RECOGNITION (2018) [15]:
 - Utiliza modelos não supervisionados (***autoencoder***) para tentar remediar a **escassez** de dados para tarefas de SER;
 - Questiona a viabilidade de aprender *features* de ***datasets*** de outros **domínios** de voz e utilizá-los para treinar modelos de classificação de emoções.

Trabalhos relacionados



- DEEP: Uma arquitetura para reconhecer emoção com base no espectro sonoro da voz de falantes da língua portuguesa (2020) [21]:
 - Utiliza o *dataset* **VERBO**;
 - Arquitetura composta por **modelos especialistas** treinados de forma supervisionada para cada emoção presente na base de dados.

Trabalhos relacionados



- Brazilian portuguese emotional speech corpus analysis (2022) [22]:
 - Utiliza o *dataset* **VERBO**;
 - Constrói quatro modelos de *machine learning* para classificar as sete emoções presentes, todos com desempenho (acurácia) acima de 74%.

Trabalhos relacionados



ANO	Referência	Features			Abordagem		Arquitetura		Português		Resultado	
		Cromática	Espectral	Prosódica	Superv.	Não superv.	ML	DL	Não	Sim	Emoção	Intensidade
2007	[65]			X	X		X		X		X	
2013	[14]		X	X	X		X	X	X			
2018	[62]		X		X		X		X		X	
2018	[15]		X		X	X		X	X		X	
2020	[21]	X	X	X	X					X	X	
2021	[22]	X	X	X	X		X	X		X	X	
2023	BRAVO		X		X	X	X	X		X		X



5) BRAVO: *Brazilian Portuguese Emotional Intensity Recognition Assistant for Voice*

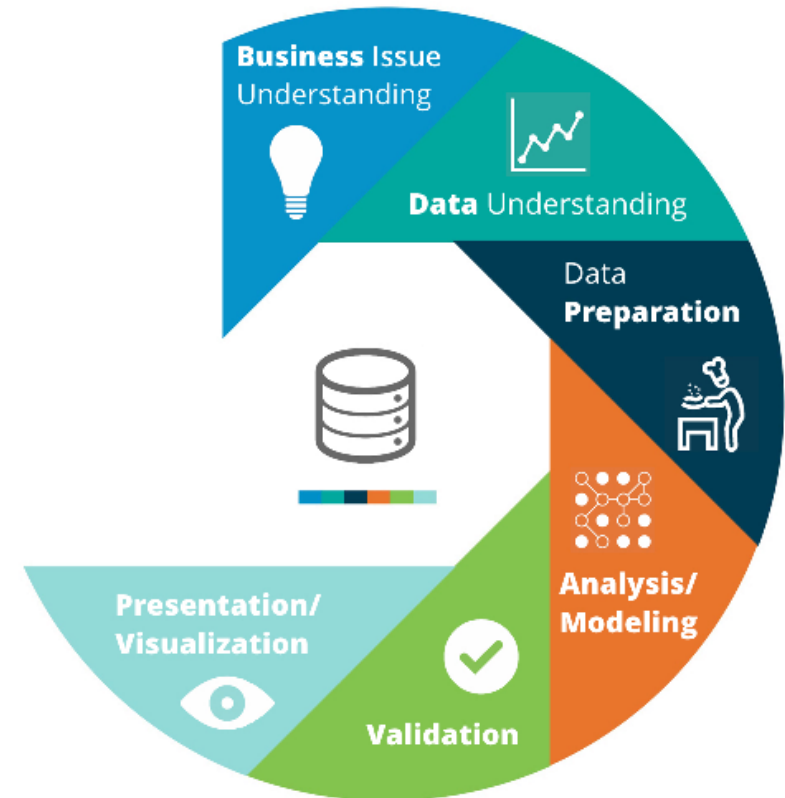
BRAVO – Ideação



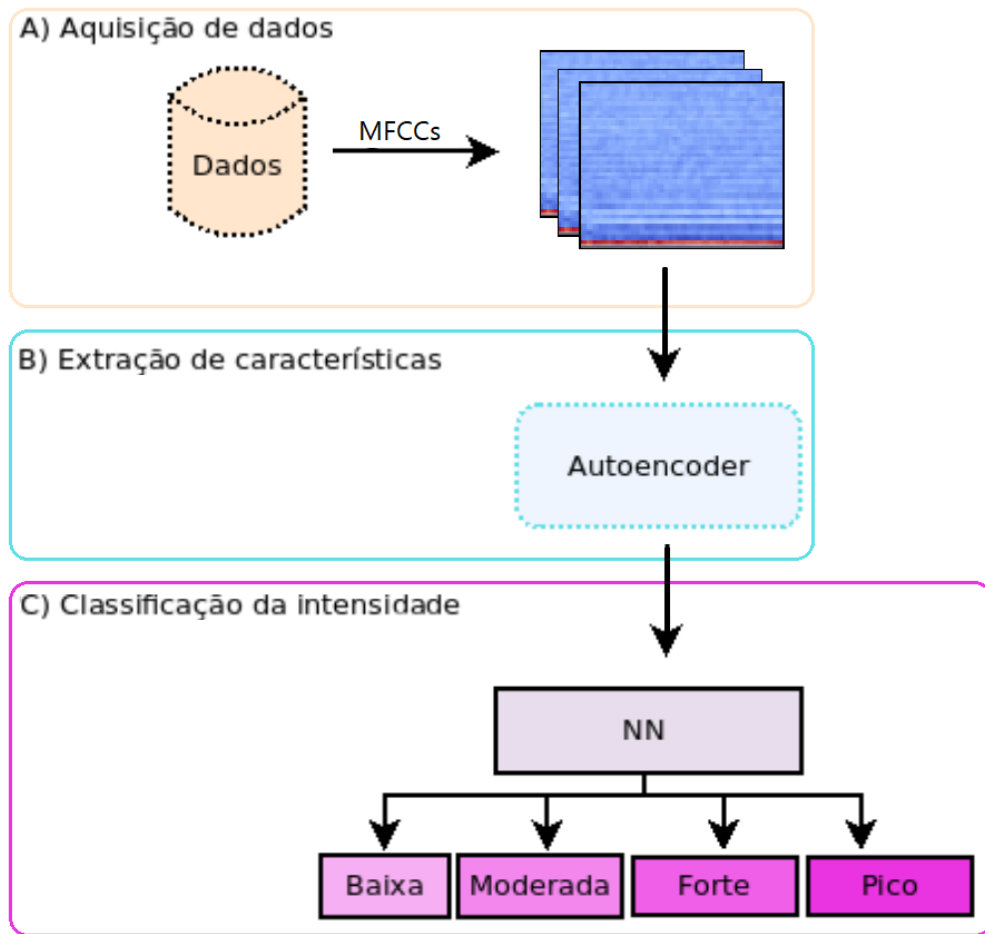
BRAVO – Metodologia



- Obtenção dos dados e consolidação do *dataset*;
- Extração de características dos dados;
- Treinamento e validação:
 - Modelo não supervisionado;
 - Modelo supervisionado;
- Decomposição dos dados;
- Análise dos resultados.



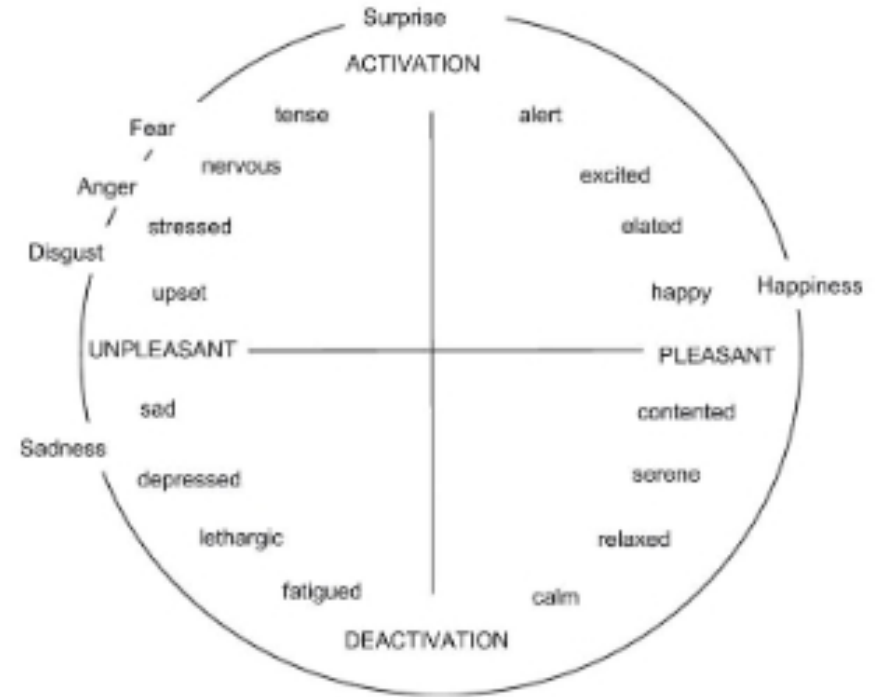
BRAVO – Visão geral



BRAVO – Emoções



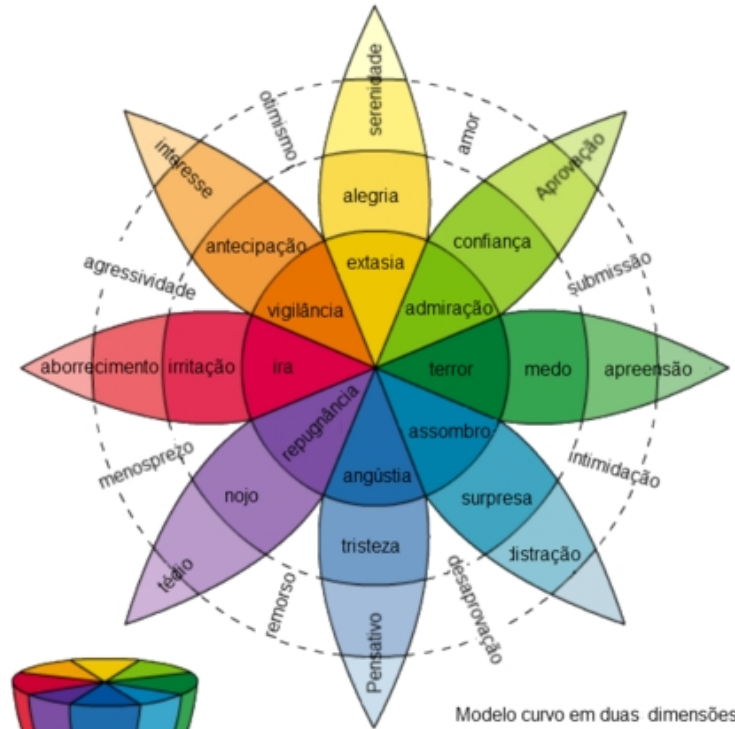
- Modelo circumplexo de Russell:
 - Valência (eixo horizontal):
 - Prazer e desprazer
 - Ativação (eixo vertical):
 - Vigor e quietude



BRAVO – Intensidade



- Modelo de Plutchik:
 - Intensidade: Sentido radial
 - Emoções compostas



BRAVO – VERBO



Distribuição por classe das 1167 sentenças do *dataset* VERBO

Classe	Total
Raiva	167
Nojo	167
Medo	166
Alegria	166
Tristeza	167
Surpresa	167
Neutro	167

BRAVO – VIVAE



Distribuição por classe das 1085 sentenças do *dataset* VIVAE

Sentimento	Intensidade	Quantidade	Total	Sentimento	Intensidade	Quantidade	Total
Conquista	Baixa	43	161	Dor	Baixa	42	185
	Moderada	40			Moderada	47	
	Pico	39			Pico	45	
	Forte	39			Forte	51	
Raiva	Baixa	42	174	Prazer	Baixa	42	202
	Moderada	44			Moderada	54	
	Pico	44			Pico	52	
	Forte	44			Forte	54	
Medo	Baixa	42	176	Surpresa	Baixa	51	187
	Moderada	43			Moderada	41	
	Pico	46			Pico	46	
	Forte	45			Forte	49	

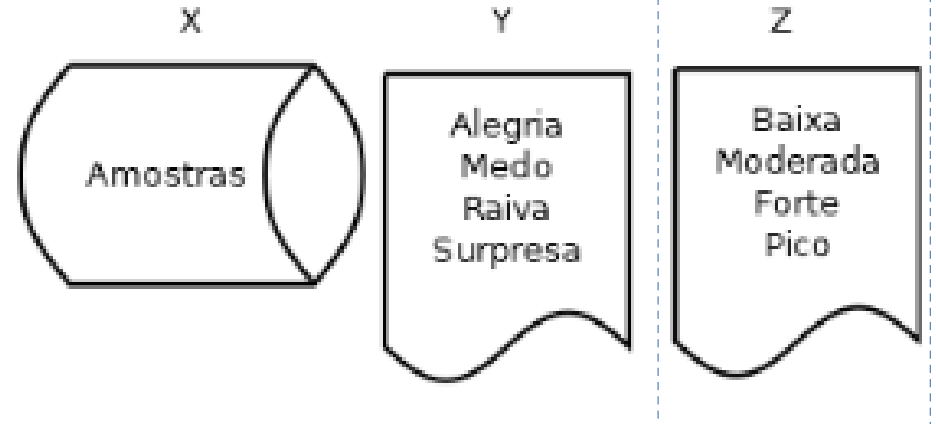
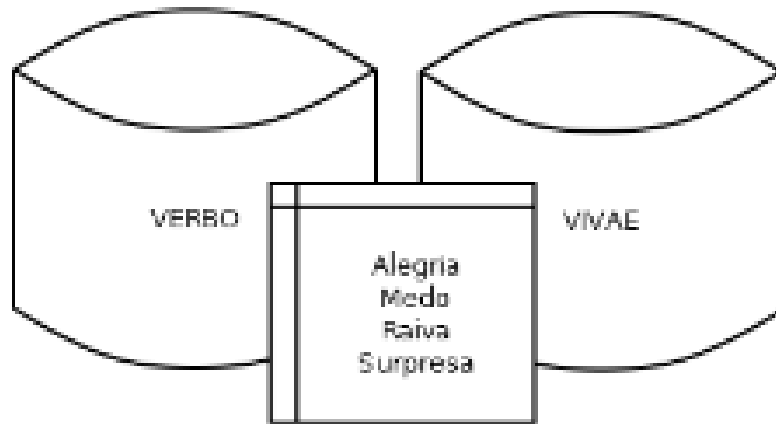
BRAVO – VIVAE



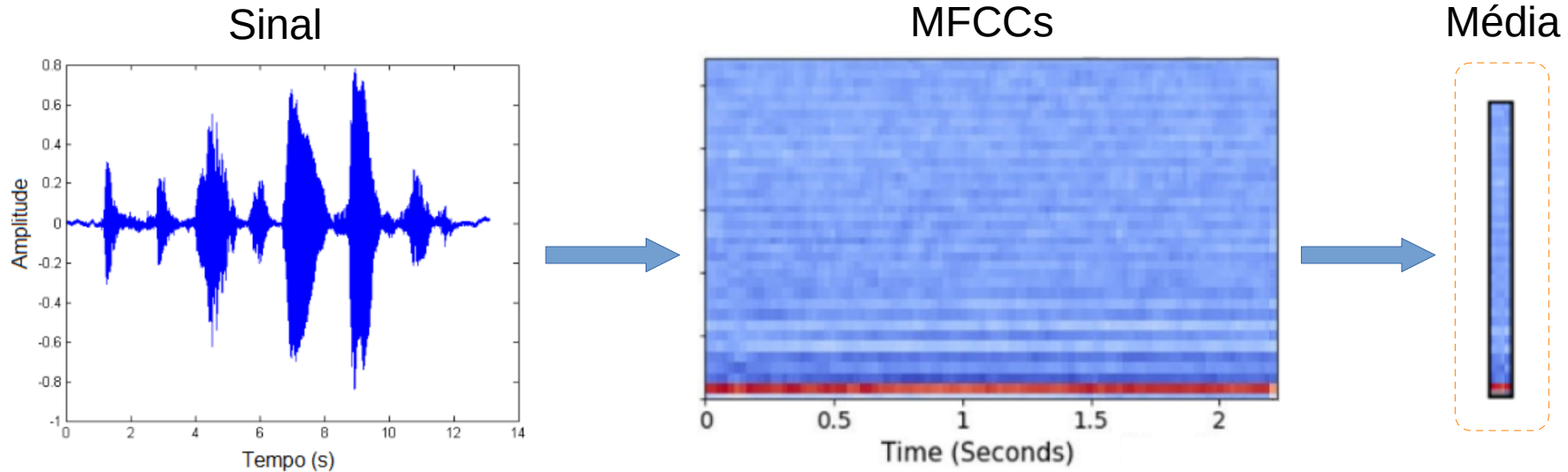
Distribuição por intensidade das 1085 sentenças do *dataset* VIVAE

Intensidade	Total
Baixa	262
Moderada	269
Forte	272
Pico	282

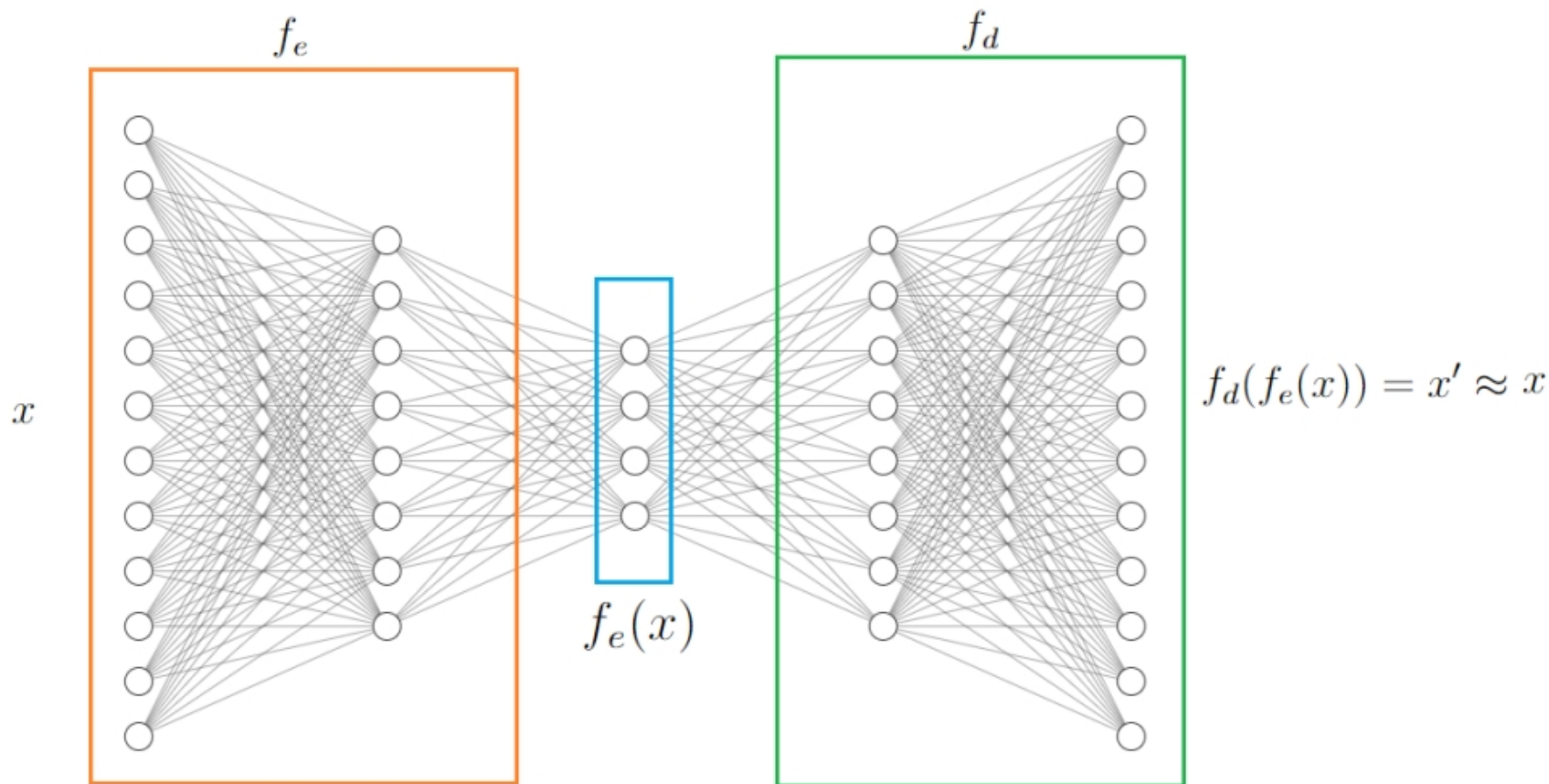
BRAVO – VERBO e VIVAE



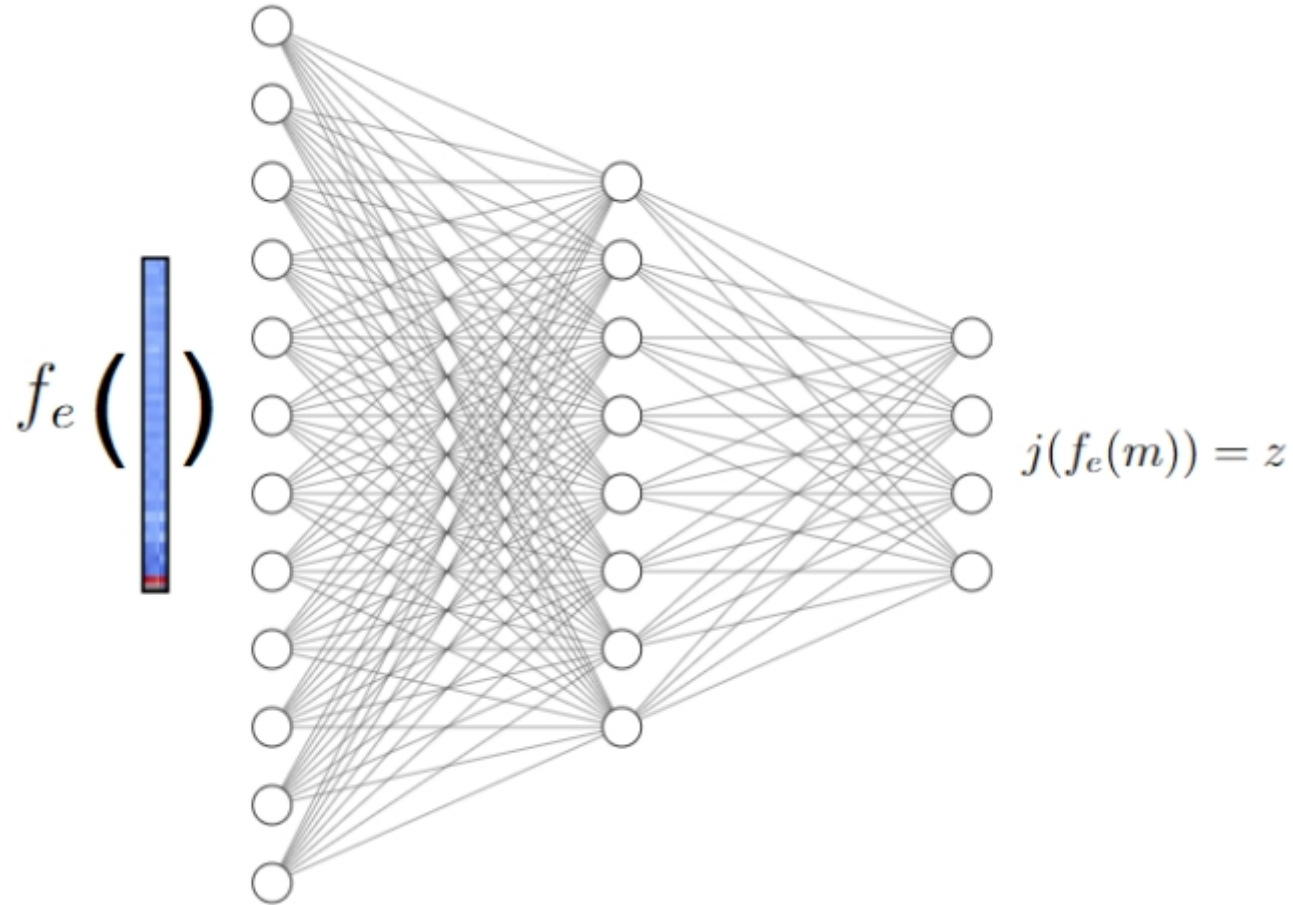
BRAVO – Extração de características



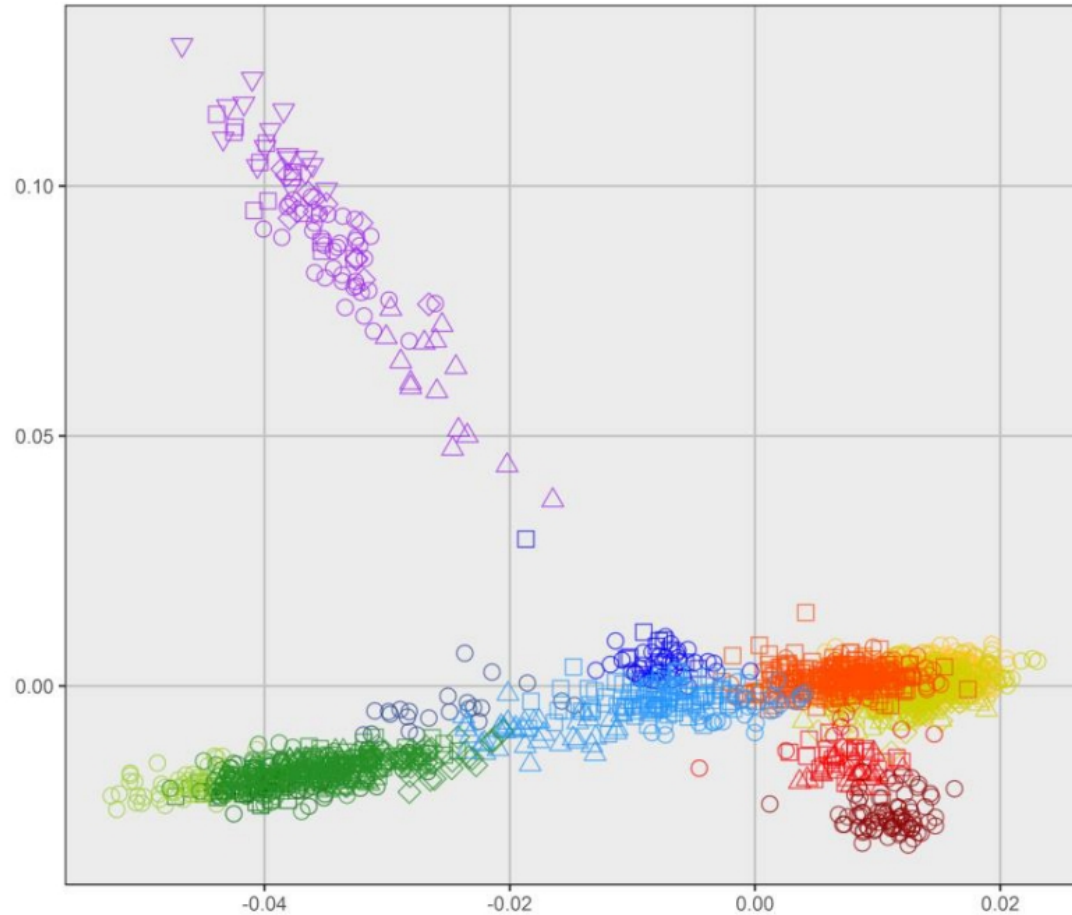
BRAVO – Não Supervisionado



BRAVO – Supervisionado



BRAVO – Decomposição





6) Cronograma

Cronograma



- Atividades:
 - 1) Obtenção dos créditos exigidos pelo programa de mestrado;
 - 2) Obtenção da certificação de proficiência em idioma inglês;
 - 3) Revisão da literatura para elaboração do trabalho de pesquisa;
 - 4) Planejar, realizar e analisar os experimentos projeto de pesquisa;
 - 5) Realizar a qualificação de mestrado;
 - 6) Análise e submissão para publicação dos resultados da pesquisa;
 - 7) Revisar a dissertação;
 - 8) Defesa do mestrado.

Cronograma



- Tabela do cronograma proposto, onde “X” representa atividades concluídas e “O” representa atividades a serem desenvolvidas

Ano	Trimestre	Atividades							
		1	2	3	4	5	6	7	8
2022	1	X							
	2	X		X					
	3	X		X					
	4	X	X	X	X				
2023	1	X			X				
	2	X				X			
	3	X					O	O	
	4	X							O

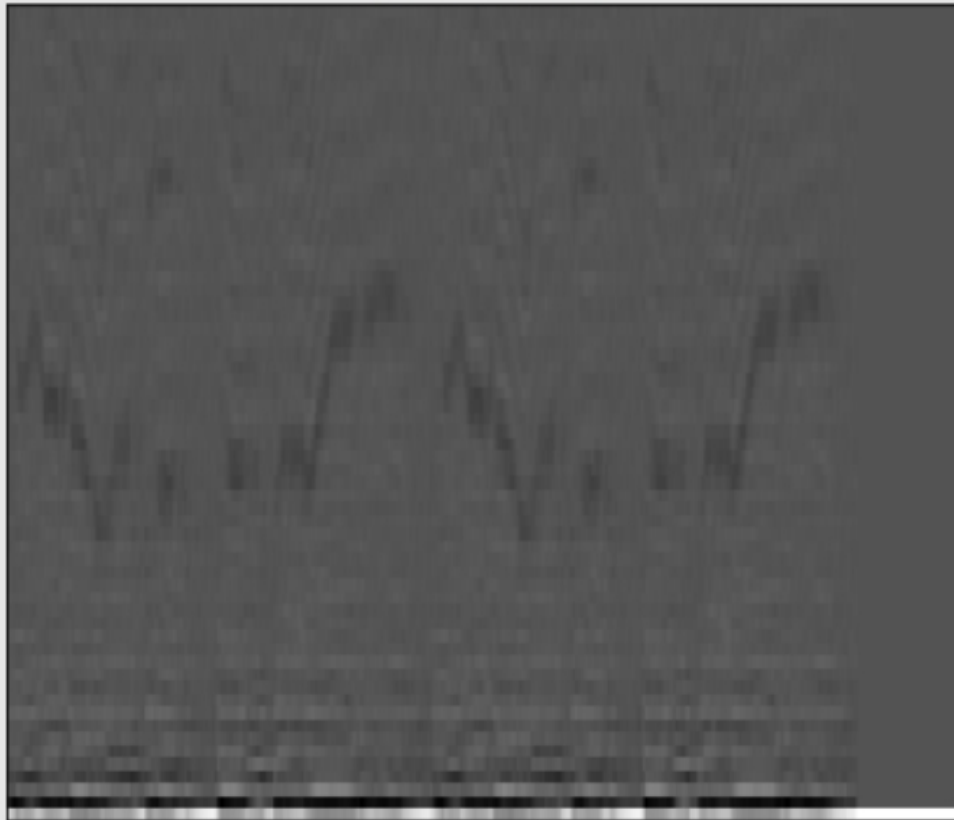


7) Resultados preliminares

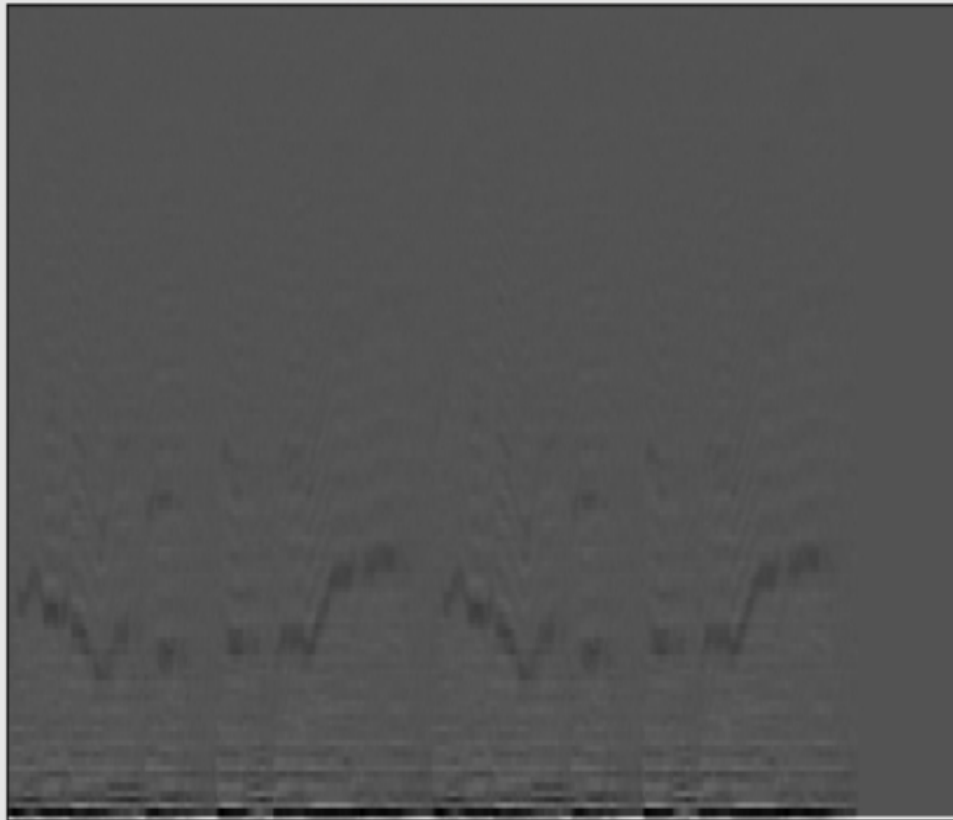
Resultados preliminares



64 MFCCs - Tile



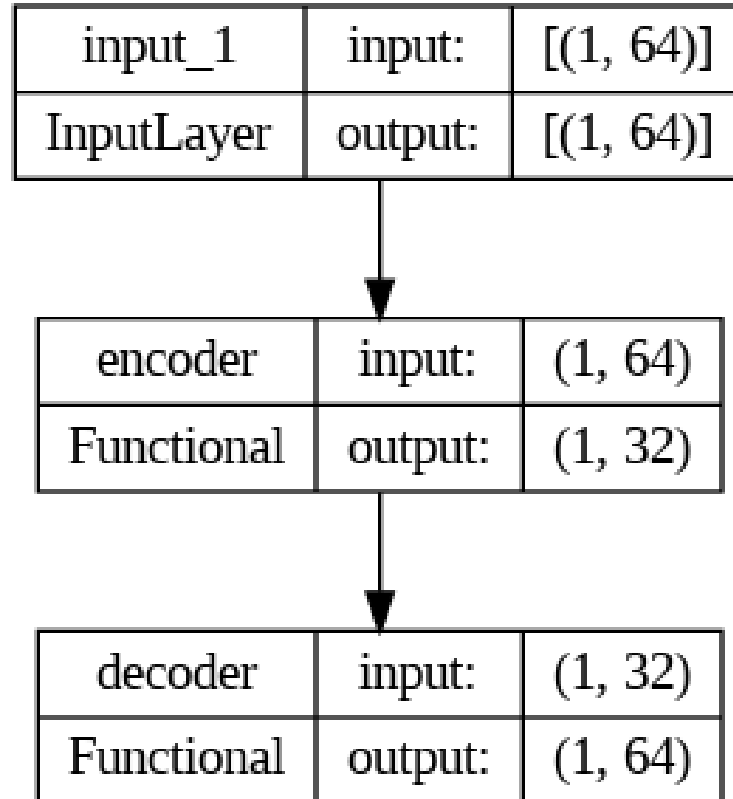
128 MFCCs - Tile





7.1) Resultados preliminares 64 MFCCs

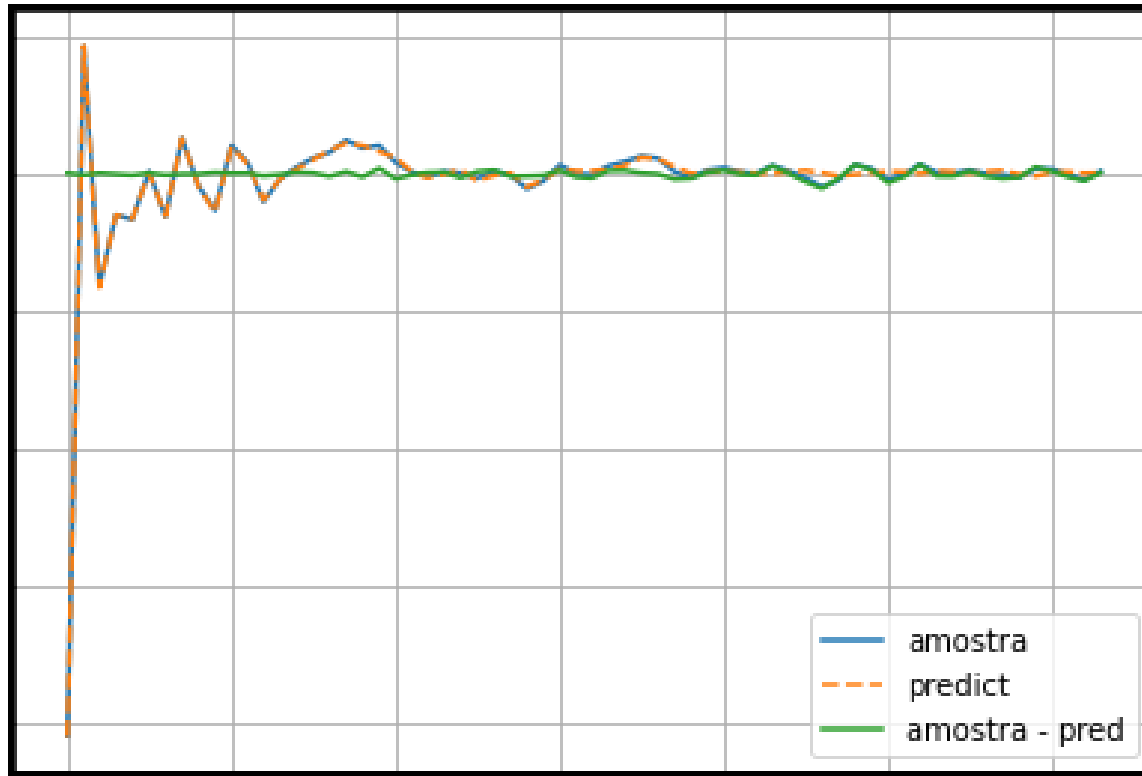
64 MFCCs - Autoencoder



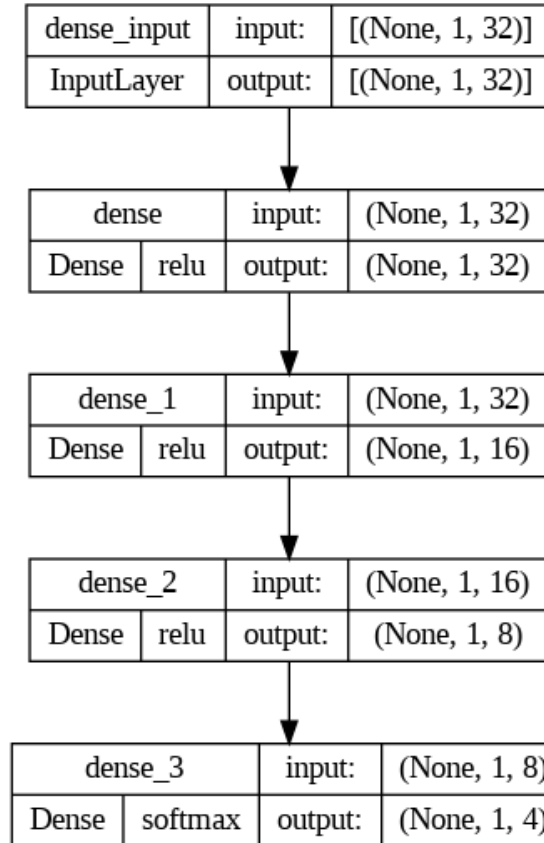
64 MFCCs - Autoencoder



$MSE = 6,40$



64 MFCCs - Classificador



64 MFCCs - Classificador

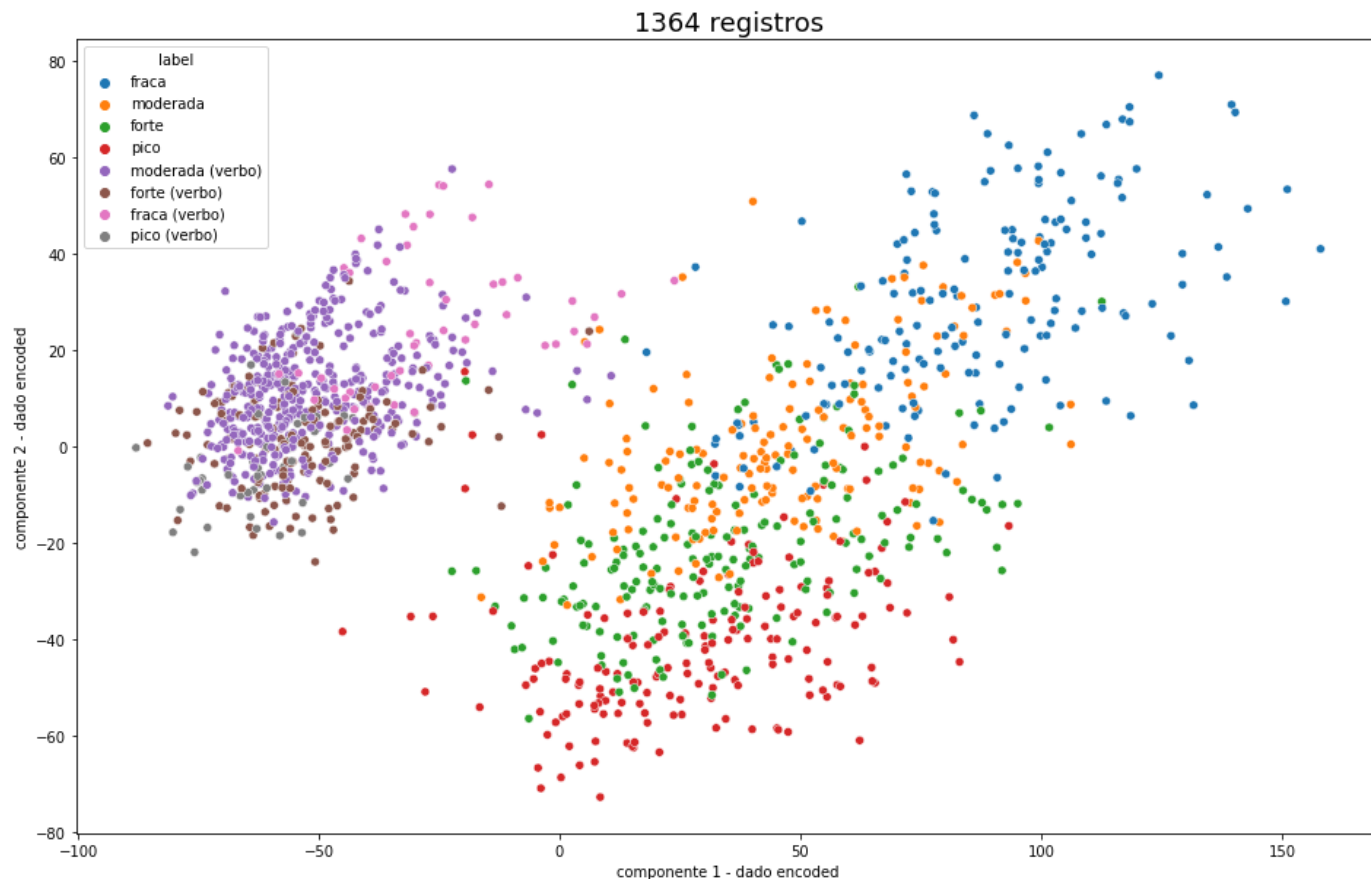


	Métricas		
Intensidade	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Fraca	0.73	0.64	0.68
Moderada	0.53	0.52	0.53
Forte	0.46	0.50	0.48
Pico	0.64	0.68	0.66

64 MFCCs - PCA



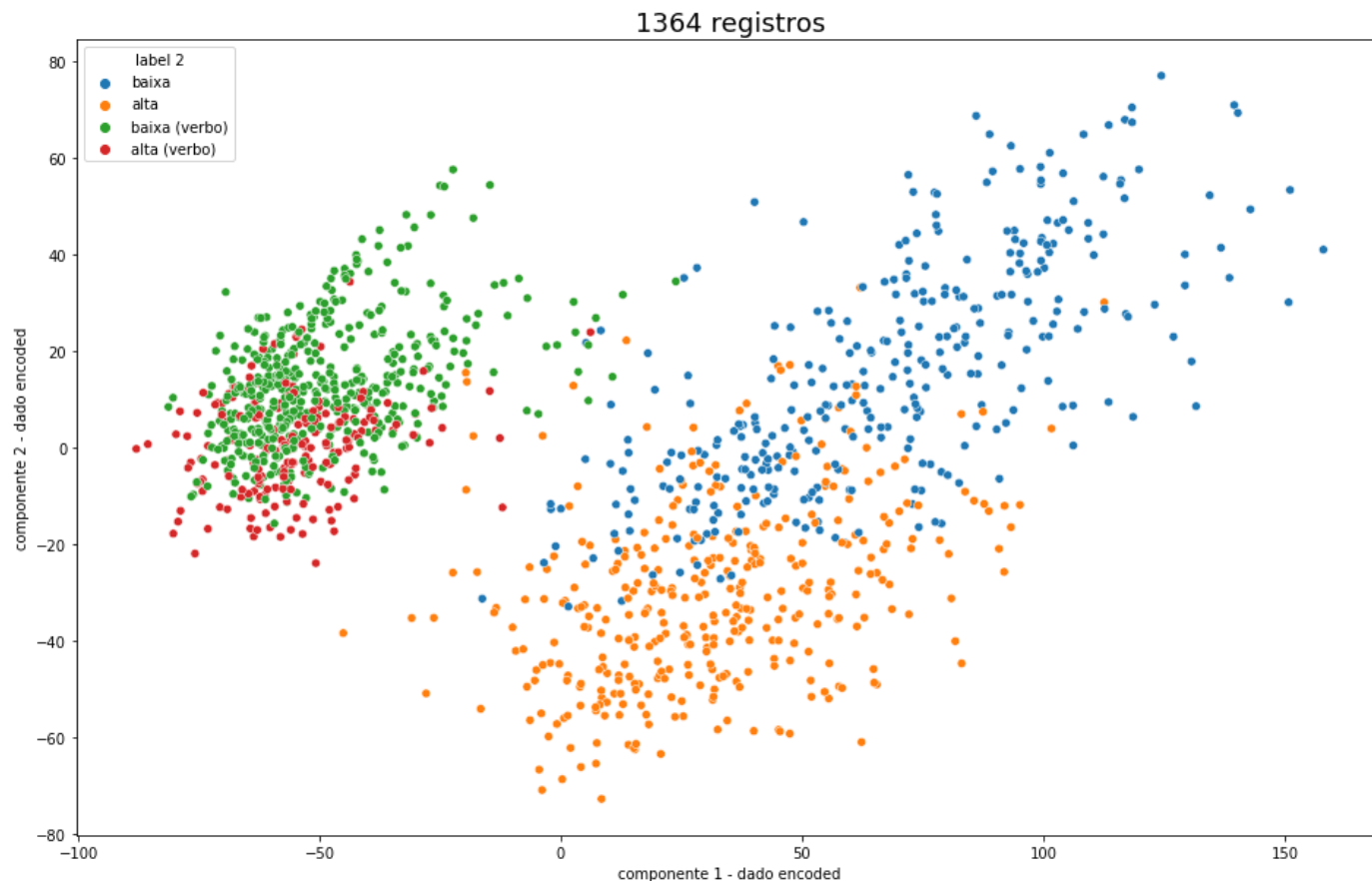
PCA aplicado ao resultado do AE em VERBO e VIVAE



64 MFCCs - PCA



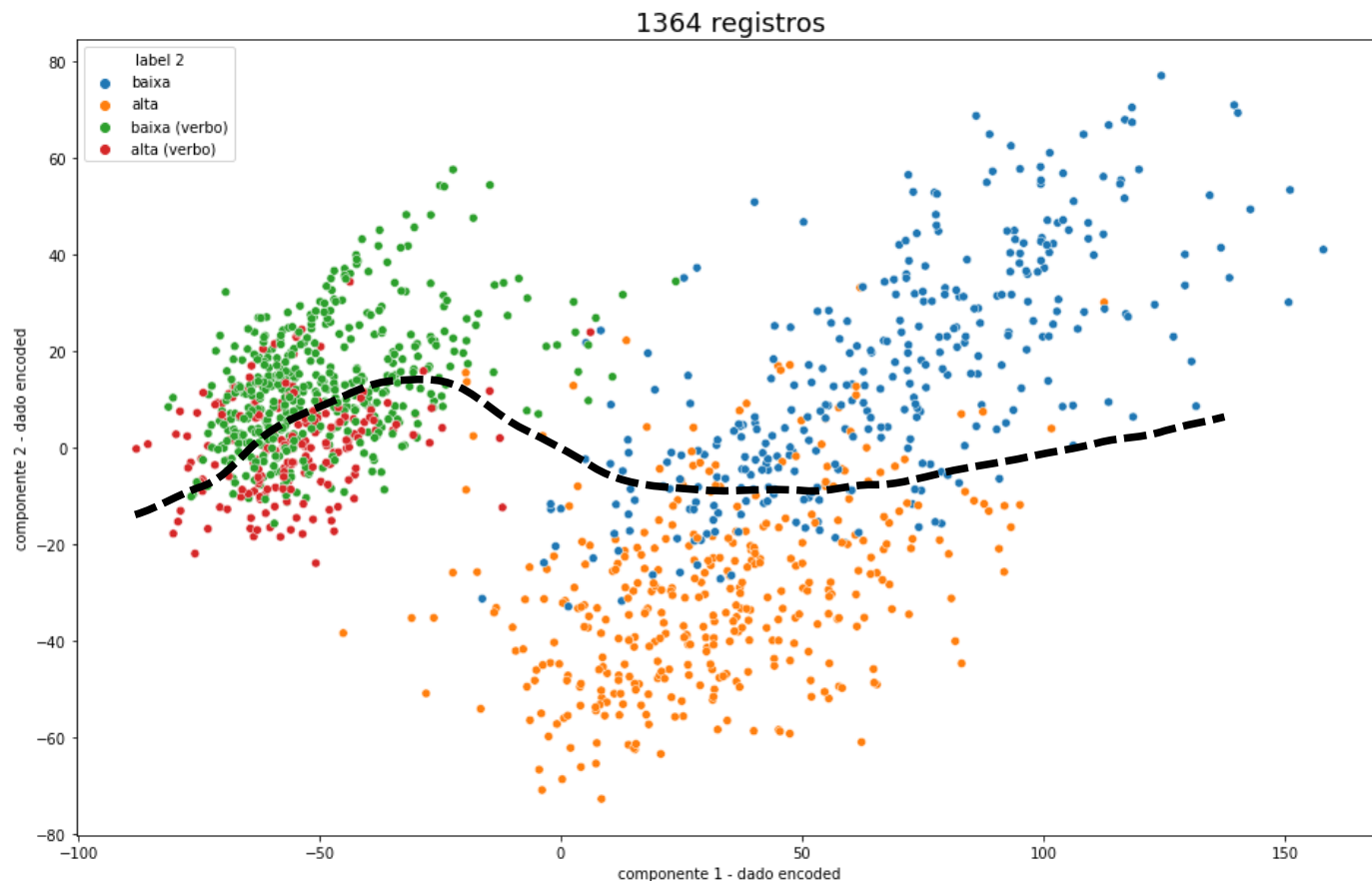
PCA aplicado ao resultado do AE em VERBO e VIVAE - 2 classes



64 MFCCs - PCA



PCA aplicado ao resultado do AE em VERBO e VIVAE - 2 classes





7.2) Resultados preliminares 128 MFCCs

128 MFCCs - Autoencoder



input_3	input:	[(1, 128)]
InputLayer	output:	[(1, 128)]



model	input:	(1, 128)
Functional	output:	(1, 64)

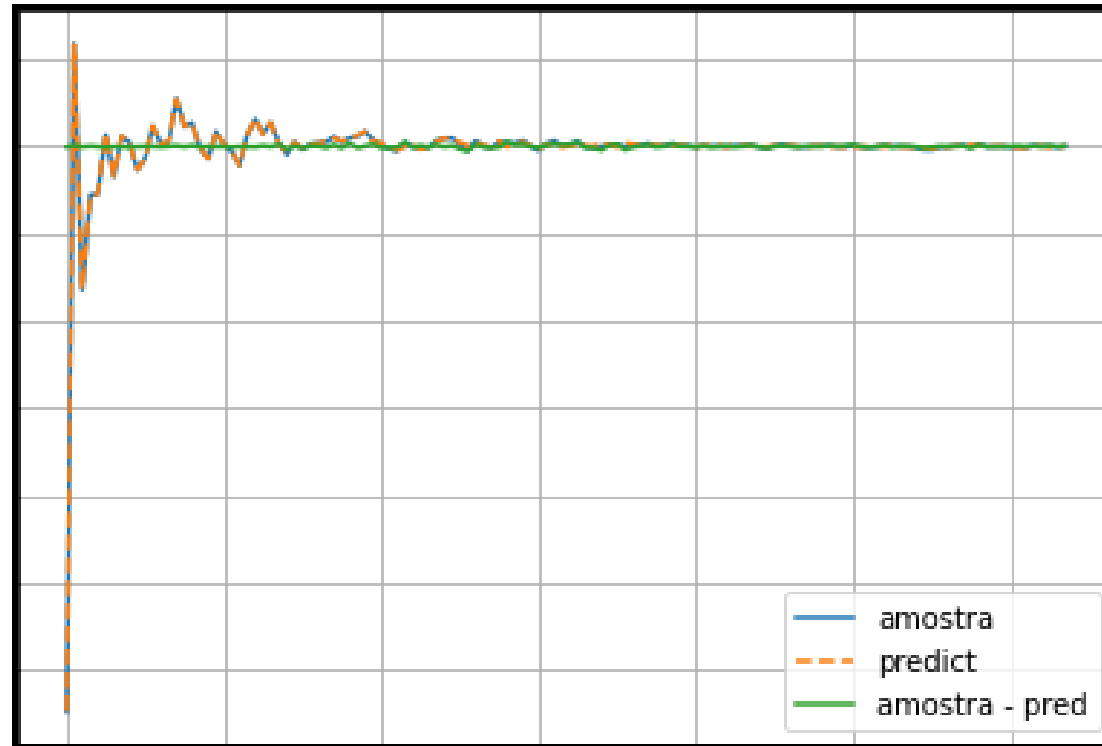


model_1	input:	(1, 64)
Functional	output:	(1, 128)

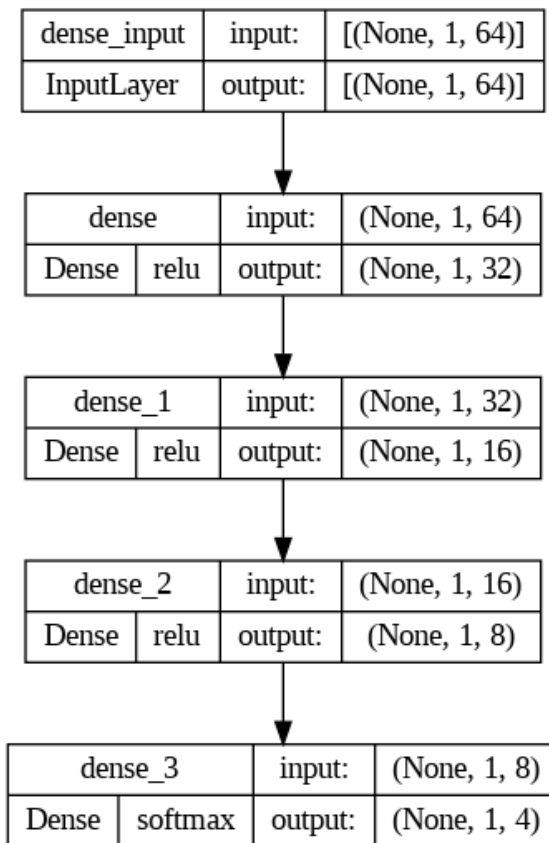
128 MFCCs - Autoencoder



$MSE = 0,84$



128 MFCCs - Classificador



128 MFCCs - Classificador

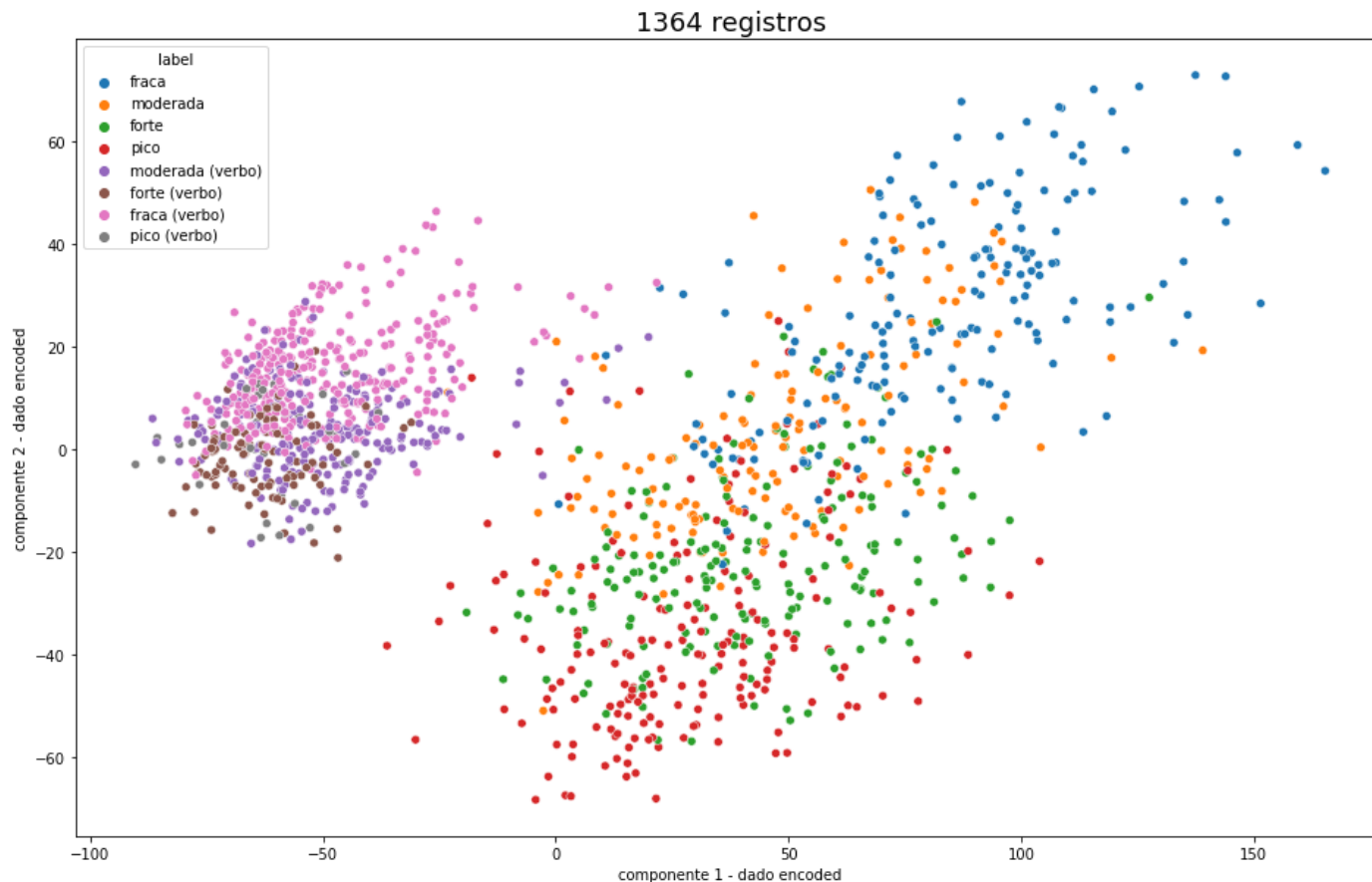


	Métricas		
Intensidade	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
Fraca	0.62	0.55	0.58
Moderada	0.55	0.48	0.51
Forte	0.49	0.50	0.49
Pico	0.57	0.73	0.64

128 MFCCs - PCA



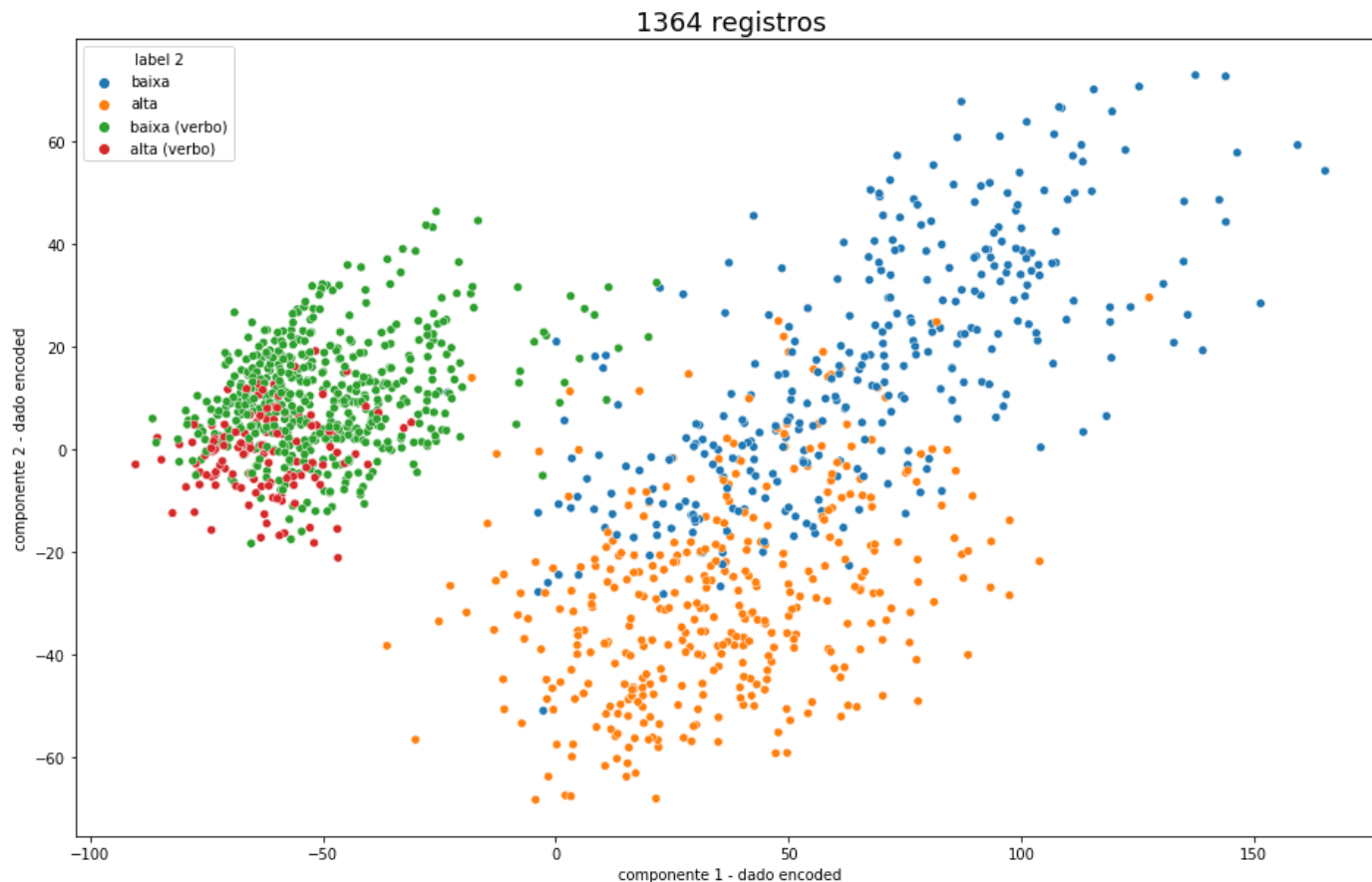
PCA aplicado ao resultado do AE em VERBO e VIVAE



128 MFCCs - PCA



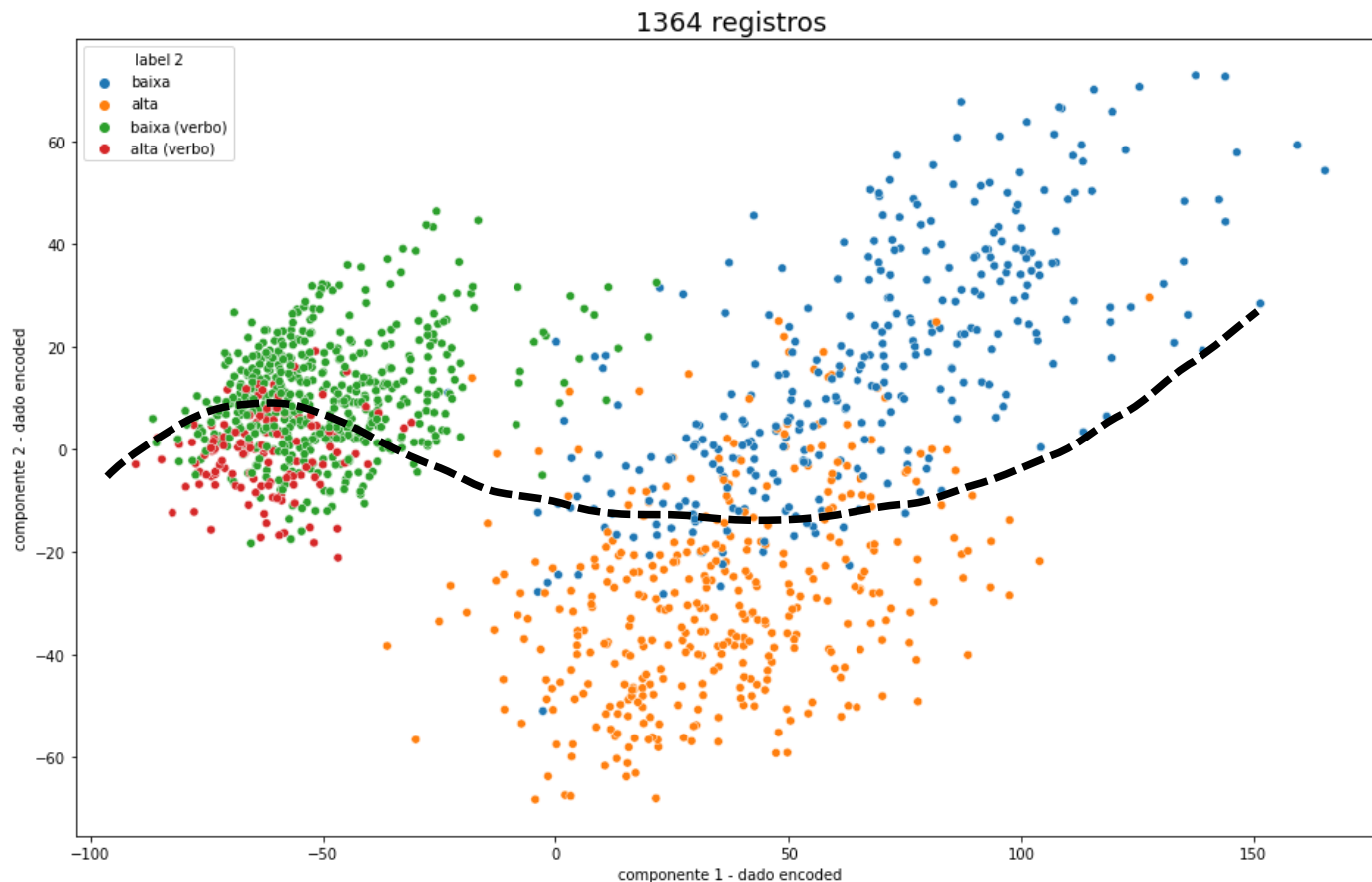
PCA aplicado ao resultado do AE em VERBO e VIVAE - 2 classes



128 MFCCs - PCA



PCA aplicado ao resultado do AE em VERBO e VIVAE - 2 classes





Obrigado!