

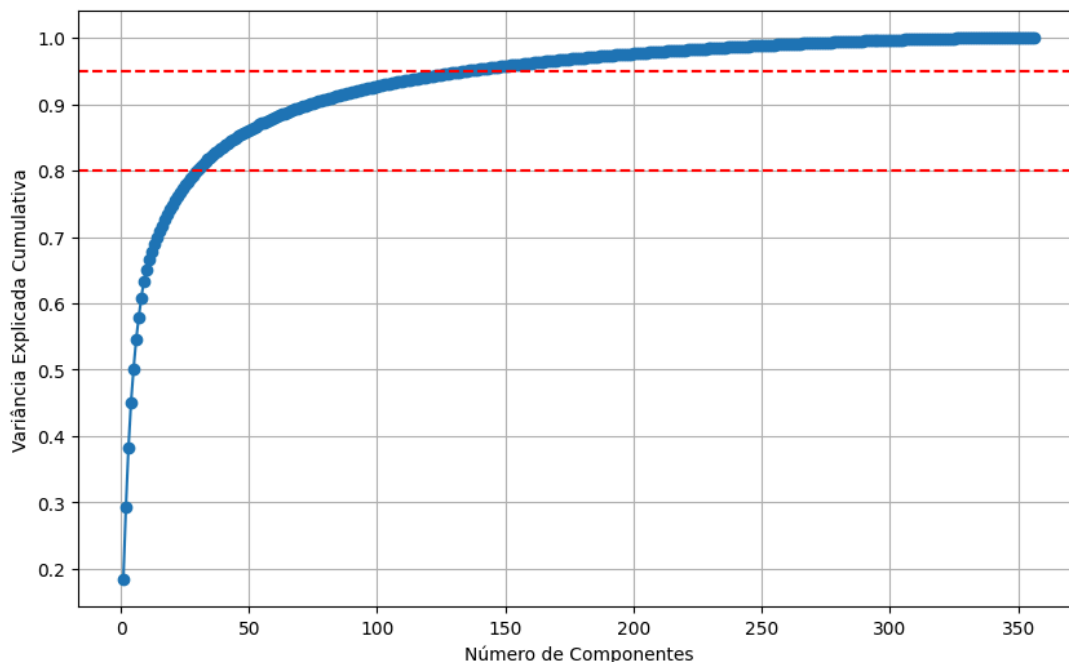
Aprendizado de Máquina I

Modelo de Reconhecimento Facial

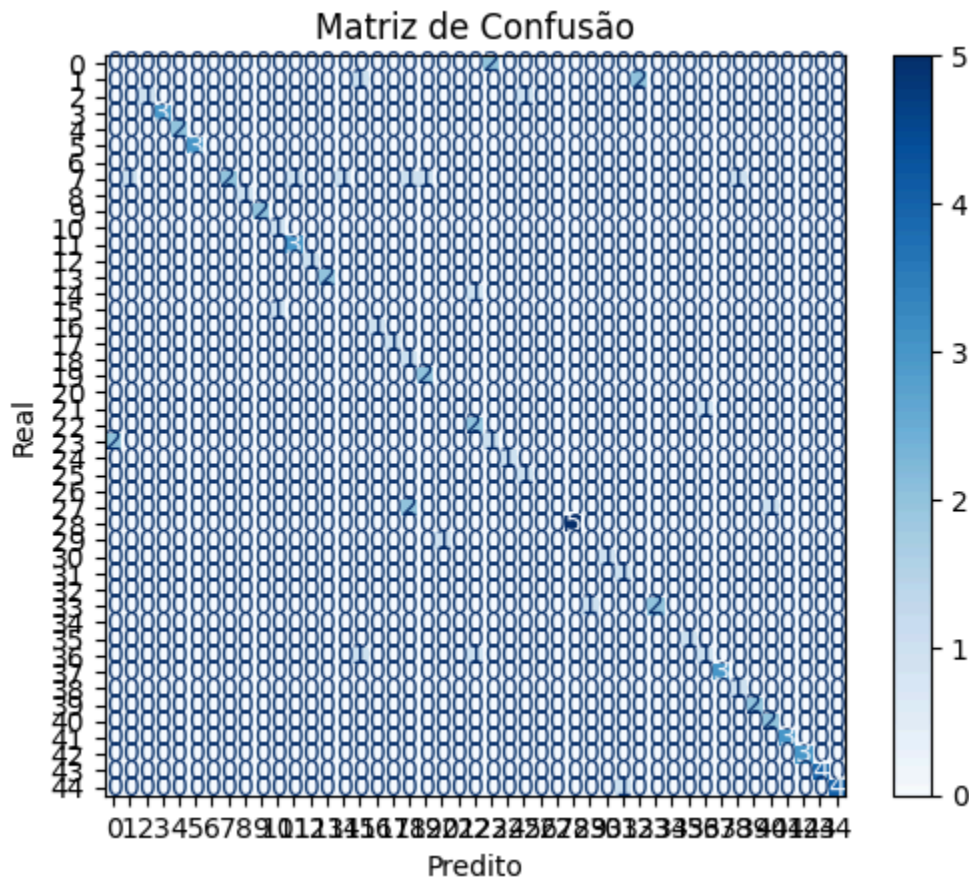
Gabriel Silvestre Mancini
Pedro Di Celio Moraes

Iniciamos removendo algumas fotografias fora do padrão definido, o dataset utilizado foi com 445 imagens, de 45 classes diferentes.

Foi utilizado o PCA randomizado pois usa algoritmos estocásticos para encontrar rapidamente uma aproximação dos primeiros componentes principais. Além disso, utilizamos `whiten = True` para normalizar os componentes principais e obter variância unitária e componentes independentes. Isso acaba sendo vantajoso para o SVM, pois componentes com muitas variâncias podem influenciar indevidamente o modelo. Com os componentes principais tendo variância unitária, temos um desempenho mais robusto do SVM. Optamos por procurar o ponto onde o PCA descrevesse 95% da variância dos dados, esse ponto foi obtido utilizando 134 componentes principais.



Para o treinamento do classificador, utilizamos uma validação cruzada através do GridSearch procurando o melhor ponto do C e do Gamma, como o kernel escolhido foi o RBF, faz sentido incluirmos o gamma na validação cruzada do GridSearch. O parâmetro C descreve a regularização e é importante que ele esteja em um ponto onde o modelo não fique muito genérico, nem muito específico, no nosso modelo o parâmetro escolhido pela validação cruzada foi $C=1000$, e o $\gamma = 0.001$.



A acurácia, o recall e o F1-Score obtidos pelo modelos foram de 0,72, 0,71 e 0,66 e observamos pela matriz de confusão que houveram algumas categorias em que o modelo teve mais dificuldade do que o normal na classificação, o que parece ser relacionado à variabilidade nas imagens carregadas pelos alunos, relacionada aos fundos e qualidades de câmeras diferentes utilizados.