**PROGRAMA DE PÓS GRADUAÇÃO EM MÉTODOS NUMÉRICOS EM ENGENHARIA**

**TÓPICOS EM VISUALIZAÇÃO CIENTÍFICA**

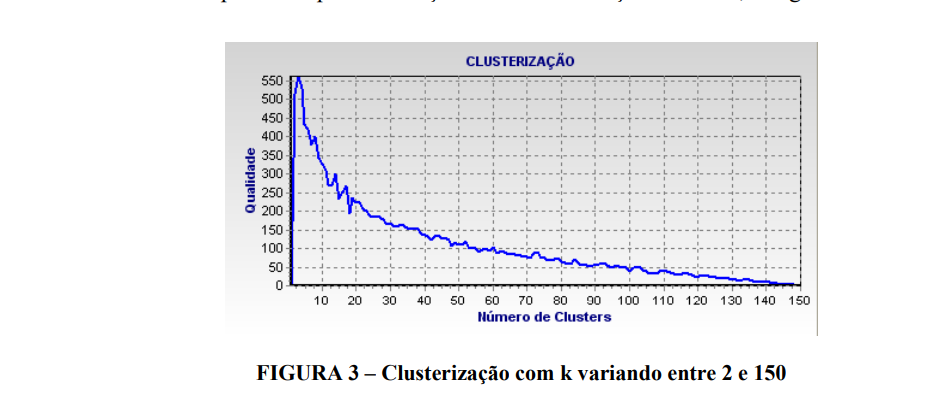
**Prof. Dr. Paulo Henrique Siqueira**

**TRABALHO FINAL**

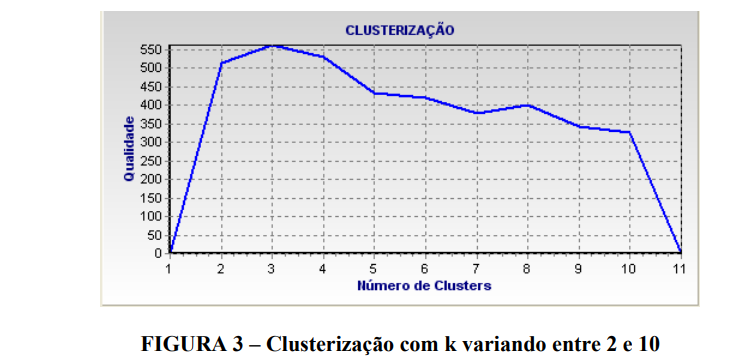
**Discente: Douglas Ferreira**

Atividade final 1 proposta da disciplina Tópicos em Visualização científica tem como objetivo propor uma forma de visualização diferente da adotada pelo autor do trabalho científico, utilizando o mesmo banco de dados. Para esta atividade foi adotado como referência o trabalho “Clusterização Automática na Redução da Dimensionalidade dos Dados”, escrito em 2008 por Éldman de Oliveira Nunes.

No trabalho o autor utiliza a técnica de clusterização para fazer a redução de da dimensionalidade de dados da base *Iris database* para fazer uma classificação das espécies da planta. De modo geral o autor explana sobre a técnica de cluster e faz um experimento no qual foi retirando variáveis do banco de dados *Iris database* e aplicando a técnica de clusterização e analisando o comportamento do seu modelo de classificação até chegar em uma quantidade satisfatória de variáveis com um bom desempenho de classificação. Segue as imagens das etapas feitas pelo autor.

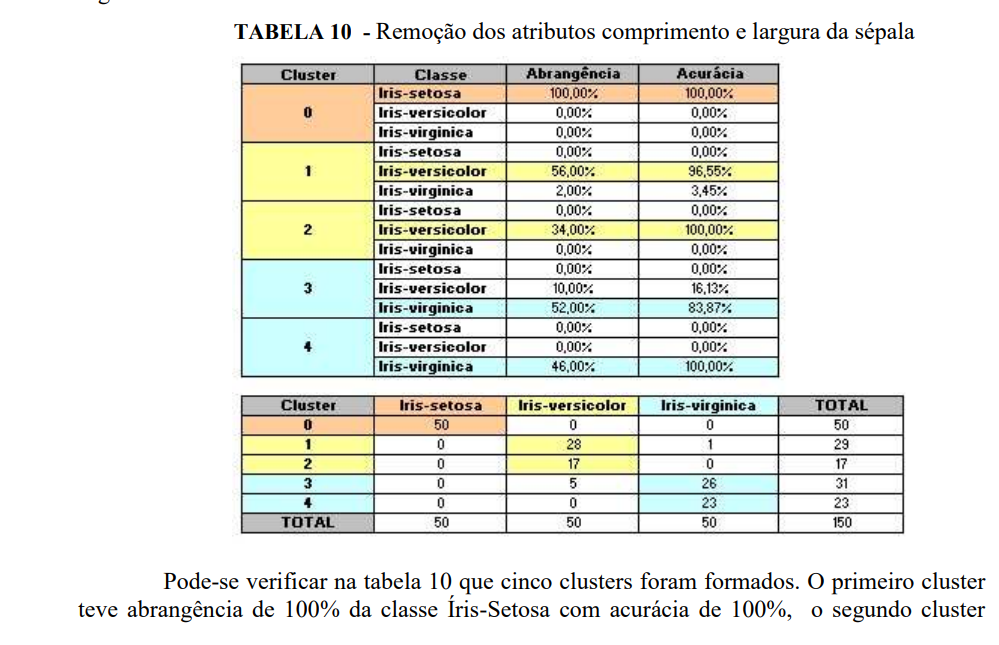


*“Clusterização Automática na Redução da Dimensionalidade dos Dados”*

**

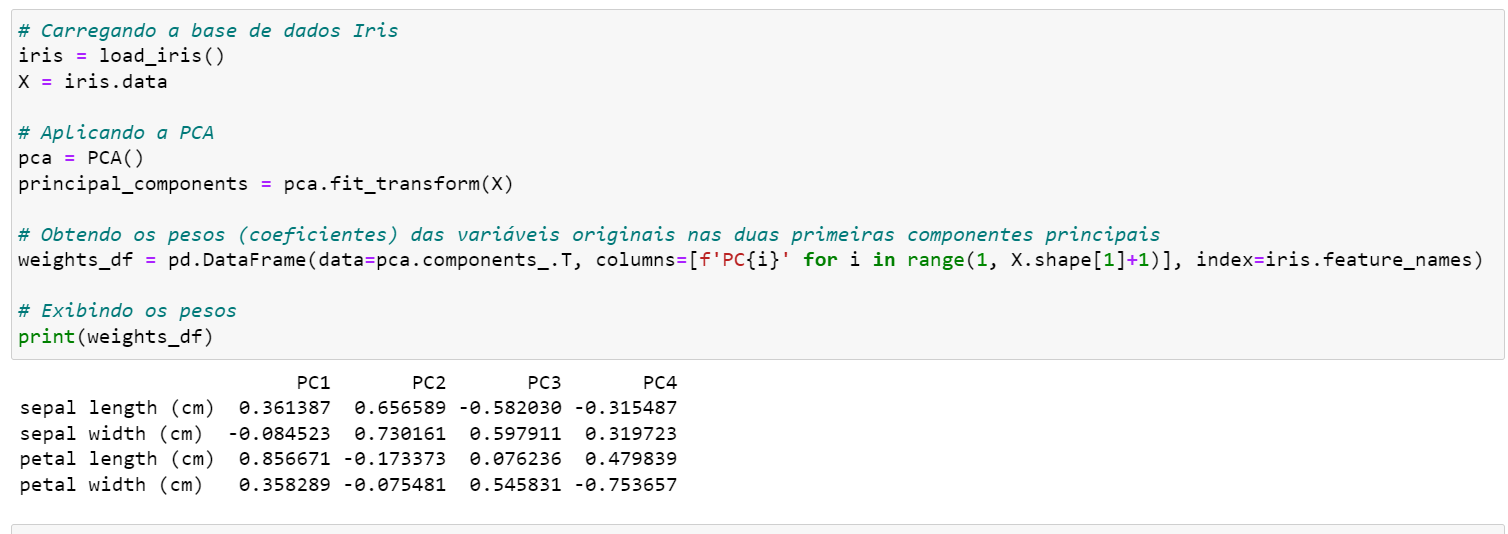
*“Clusterização Automática na Redução da Dimensionalidade dos Dados”*

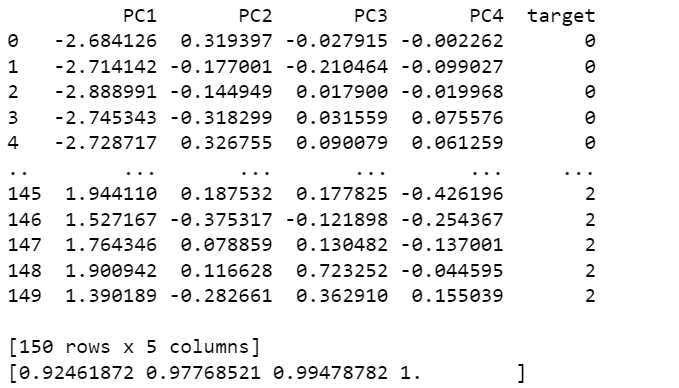
O trabalho segue com a retirada das variáveis e comparação da acurácia do modelo de classificação para verificação do desempenho do modelo, até chegar na combinação das variáveis com melhor classificação como no exemplo abaixo:

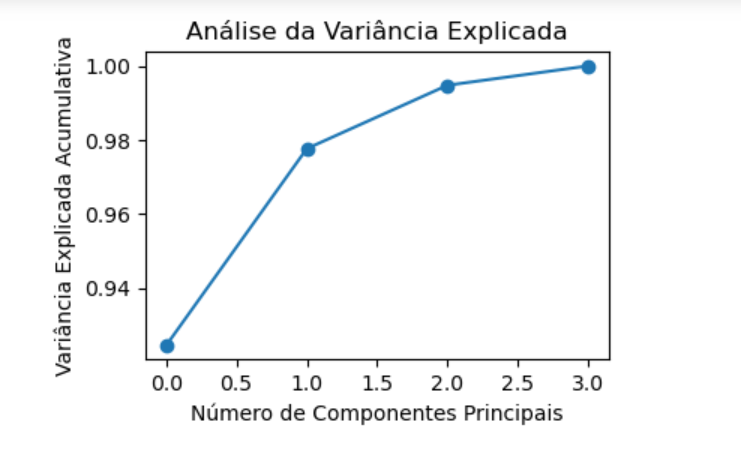
**

*“Clusterização Automática na Redução da Dimensionalidade dos Dados”*

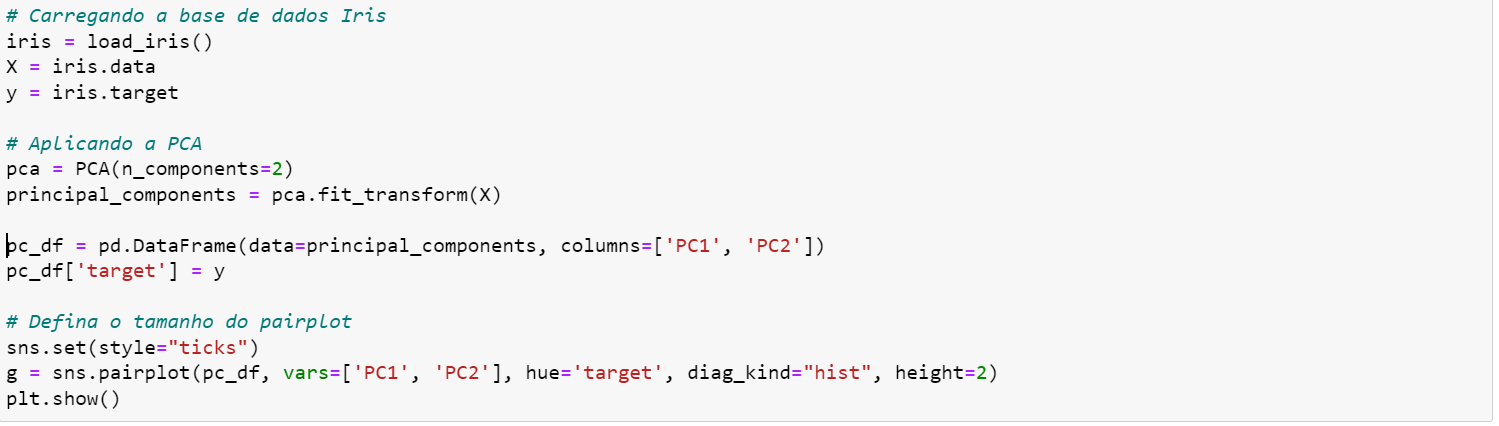
Uma outra abordagem que poderia ser feita para uma melhor seleção das variáveis seria a técnica de Analise de Componentes Principais (Pca). Tal método tem a finalidade a análise de dados visando sua redução de dimensionalidade buscando a escolha das formas representativa dos dados utilizando combinações lineares. Aplicando o método de PCA na base de dados Iris, obteremos os seguintes resultados

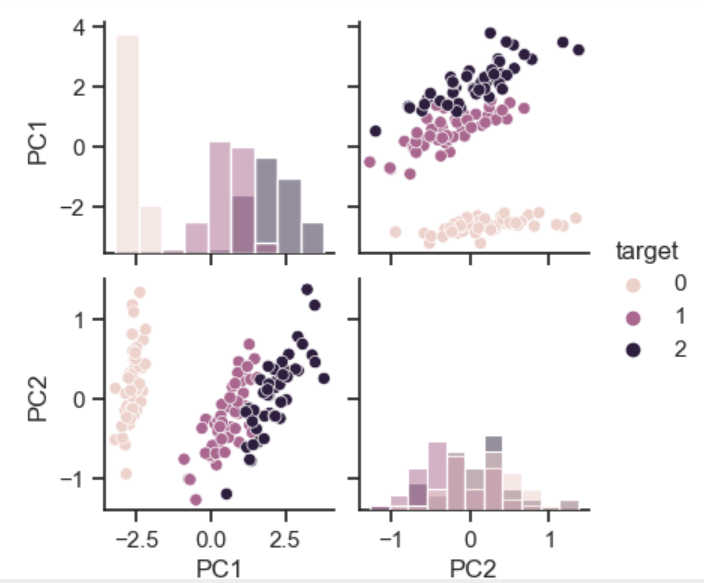


**

Podemos perceber e uma analise prévia que o método do PCA com apenas duas componentes principais é possível explicar 97,76% dos dados

Obtendo assim, uma boa separação das espécies como podemos observar

**

**

Referência bibliográfica:

Nunes. Éldman de Oliveira. Clusterização automática na redução da dimensionalidade dos Dados; Rio de Janeiro 06 de Agosto de 2008.