

Gene Expression Analysis - Prostate Cancer

1.1 Background

1.2 Objectives

The primary goal of this study is to build machine learning models for the prediction of prostate cancer between prostate cancer tumour cells and normal tissue. By leveraging machine learning techniques, the aim is to develop predictive models for disease classification.

2. Methodology

2.1 Data set

The dataset of interest was secondarily obtained from [[SBCB Lab \(ufrgs.br\)](#)]. However, the primary source of the dataset is found on [[GEO Accession viewer \(nih.gov\)](#)]. The original submitter-supplied dataset GSE22260 was obtained from GEO [Home - GEO - NCBI \(nih.gov\)](#), which was based on the platform of [GPL9115](#) Illumina Genome Analyzer II (*Homo sapiens*).

The data was submitted by ¹. In the experiment, transcriptome (polyA+) of 20 prostate cancer tumors and 10 matched normal tissues were sequenced using Illumina GAII platform. This was followed by bioinformatic approaches to identify prostate cancer specific aberrations which include gene fusion, alternative splicing, somatic mutation, etc. Data was already preprocessed by [[SBCB Lab \(ufrgs.br\)](#)].

3. Results

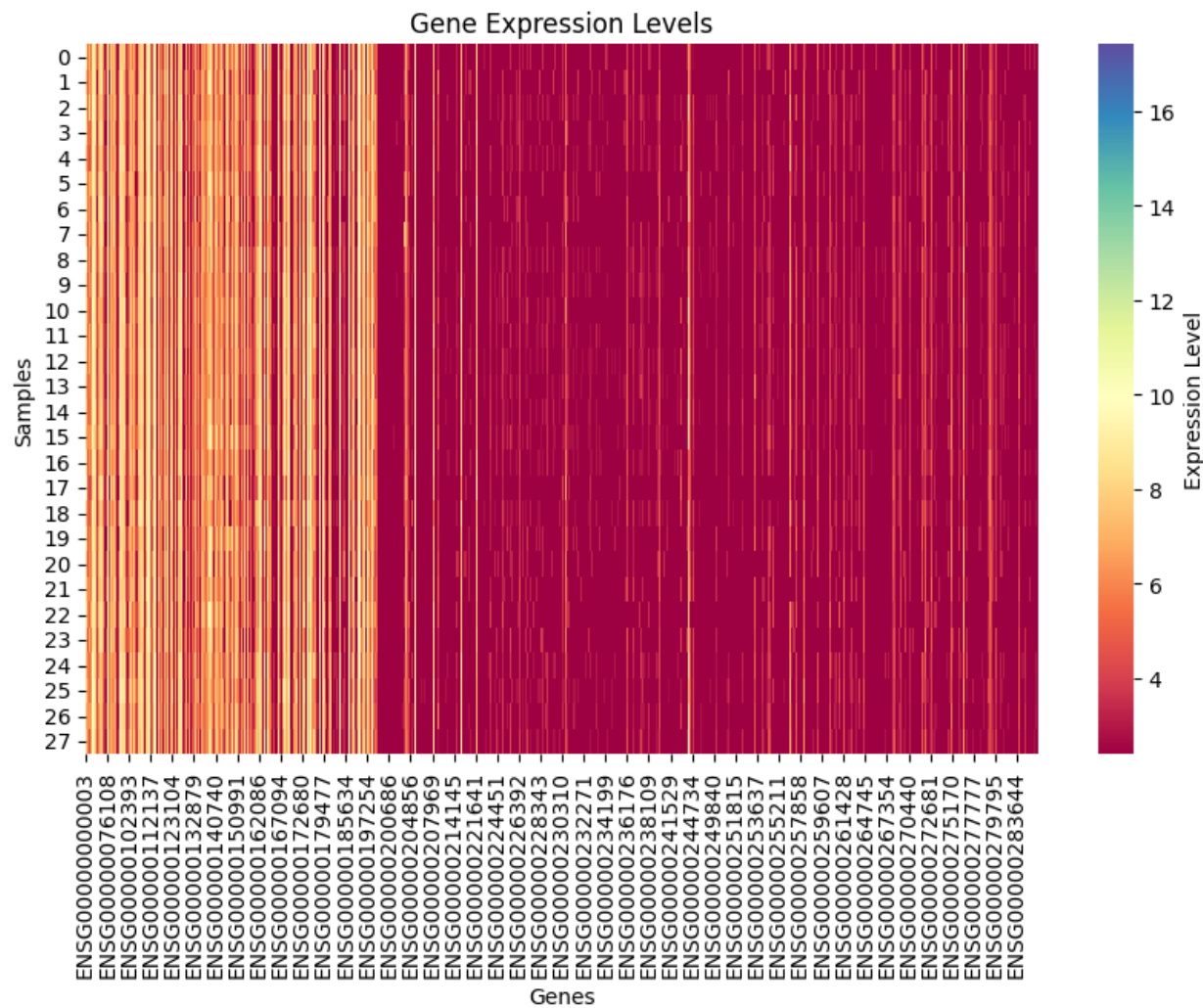


Figure 1: Gene Expression Levels Heatmap

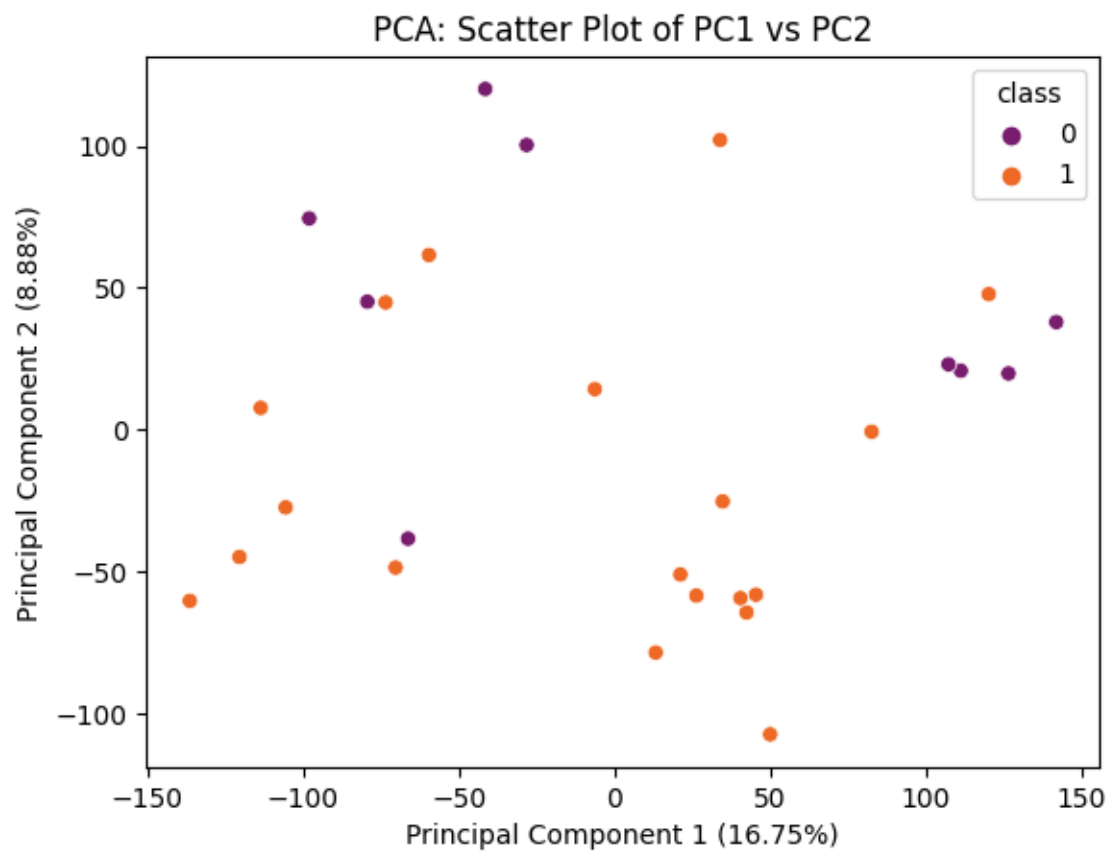


Figure 2: PCA scatter plot

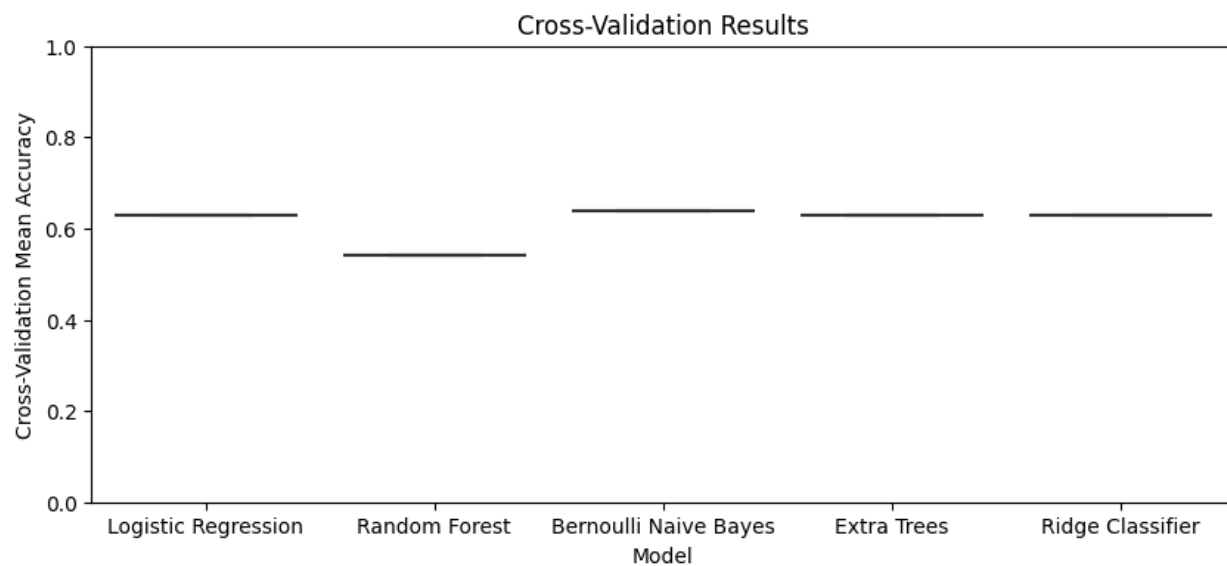


Figure 3: ML models cross-validation results

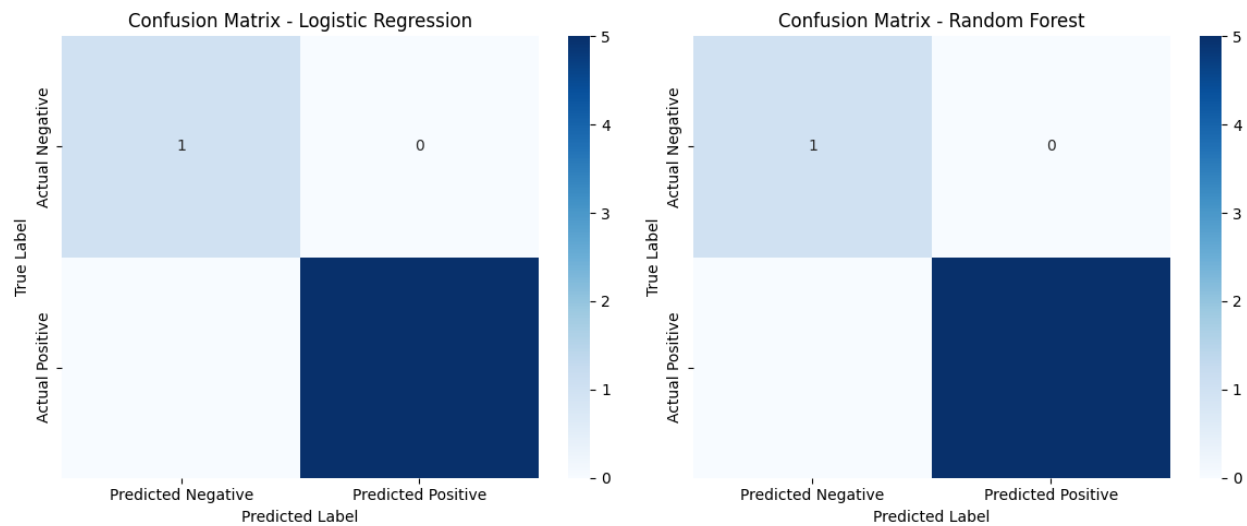


Figure 4: Confusion matrix of logistic regression and random forest models

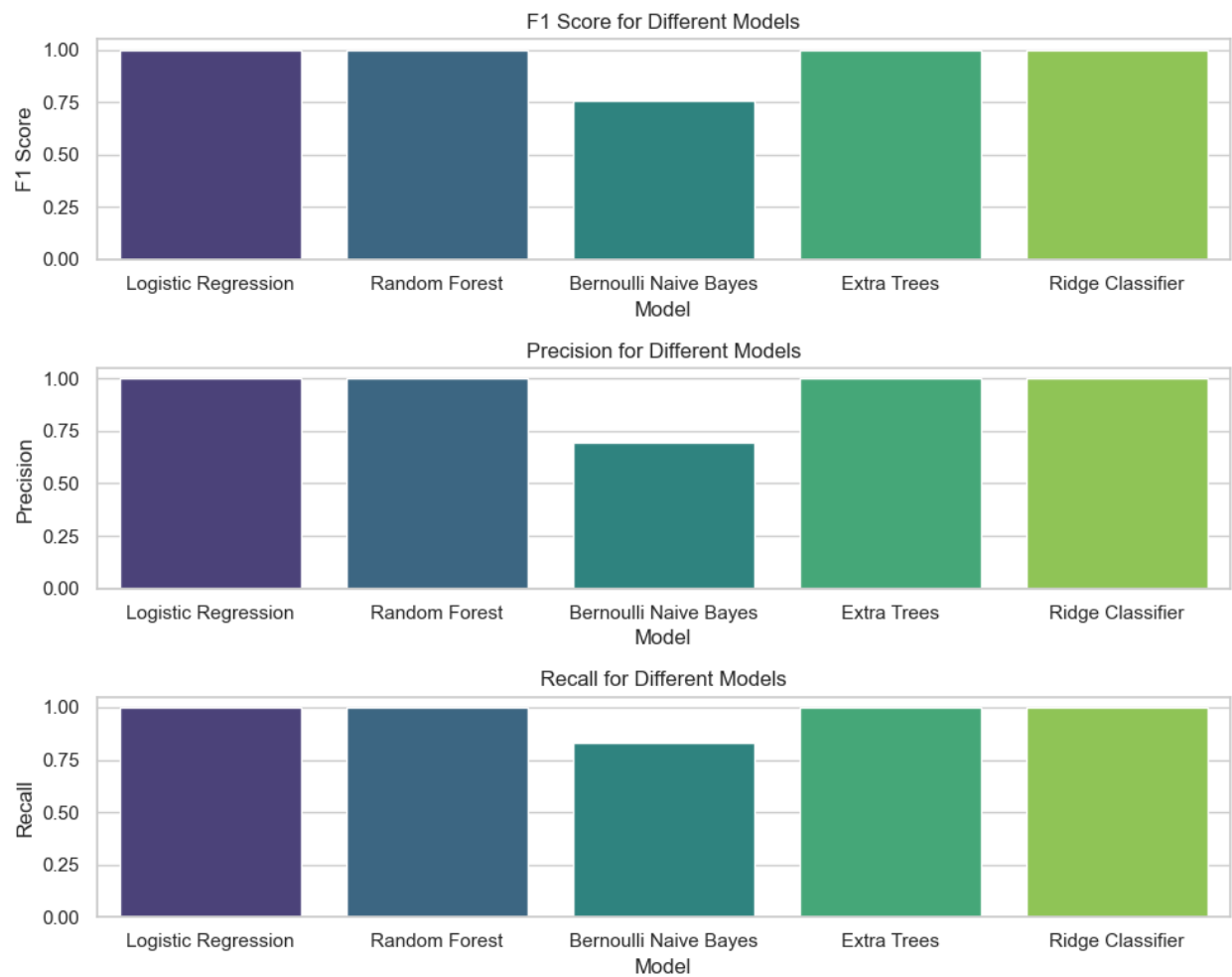


Figure 5: F1 score, precision and recall bar charts for top 5 models



Figure 6: Top 10 genes from various models

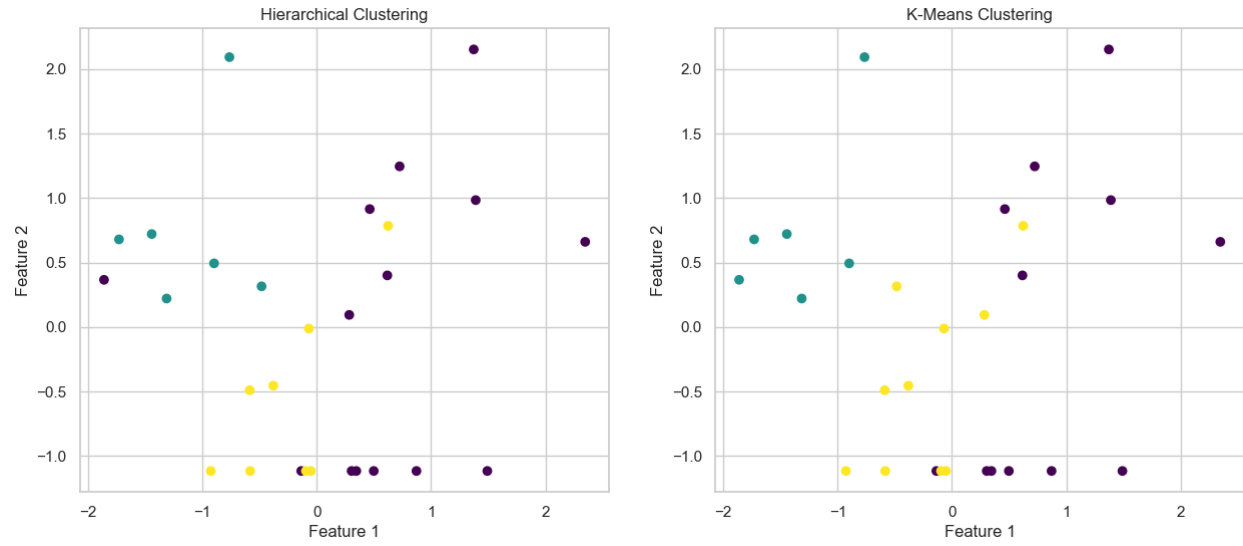


Figure 7: Scatterplot for Hierarchical and K-Means Clustering

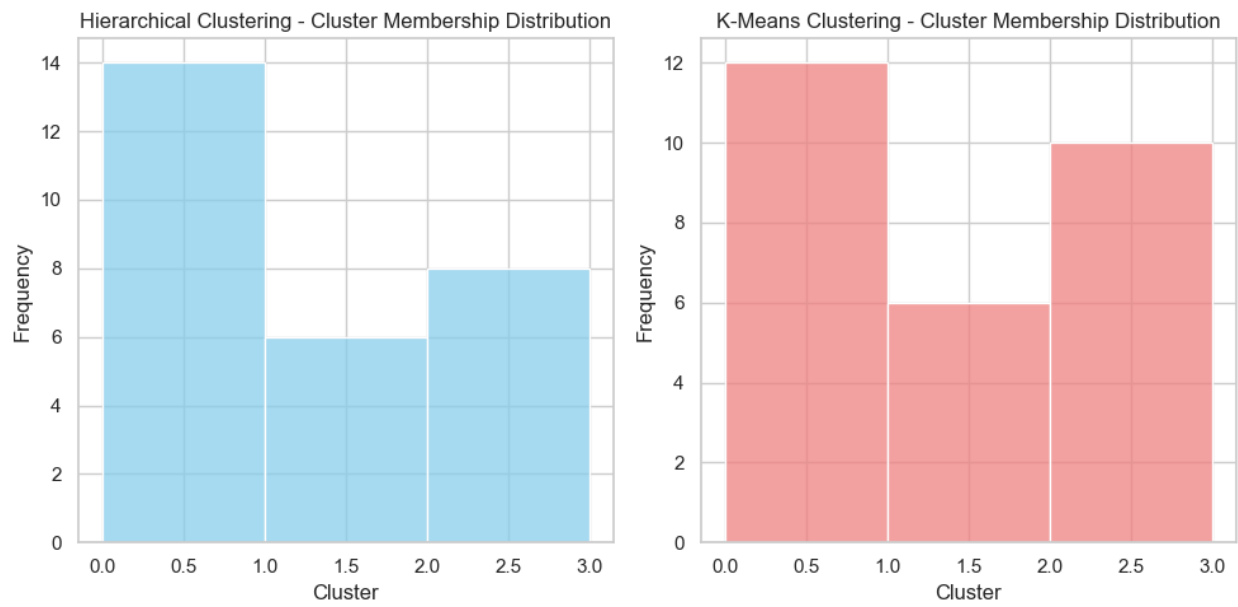


Figure 8: Cluster Membership Distributions

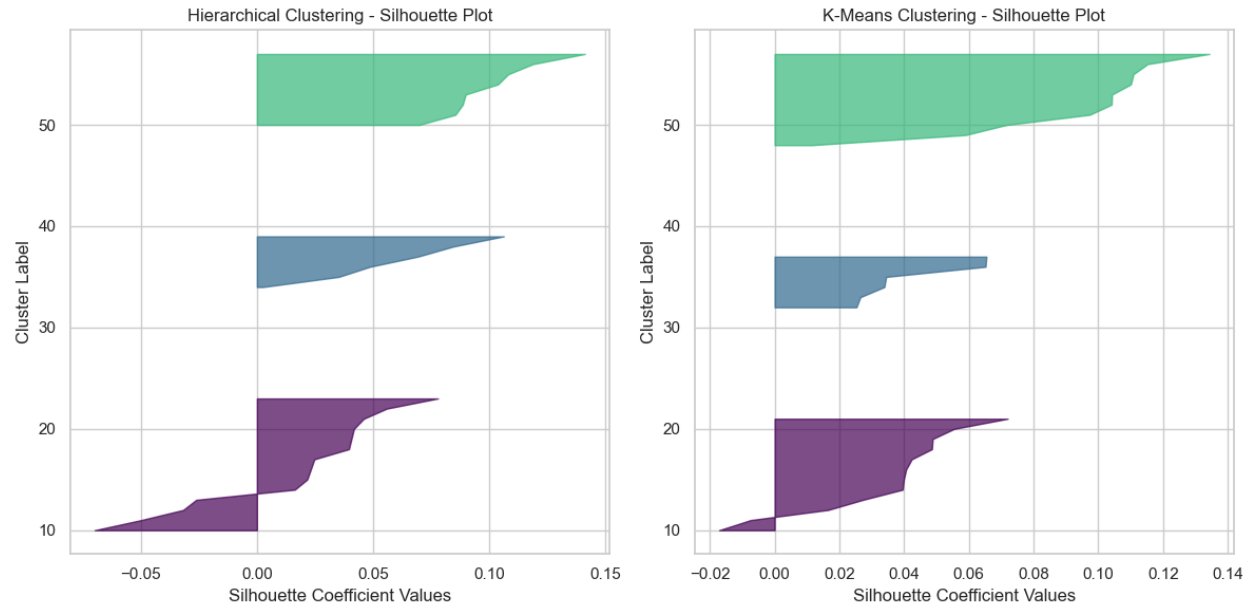


Figure 9: Silhouette Plot

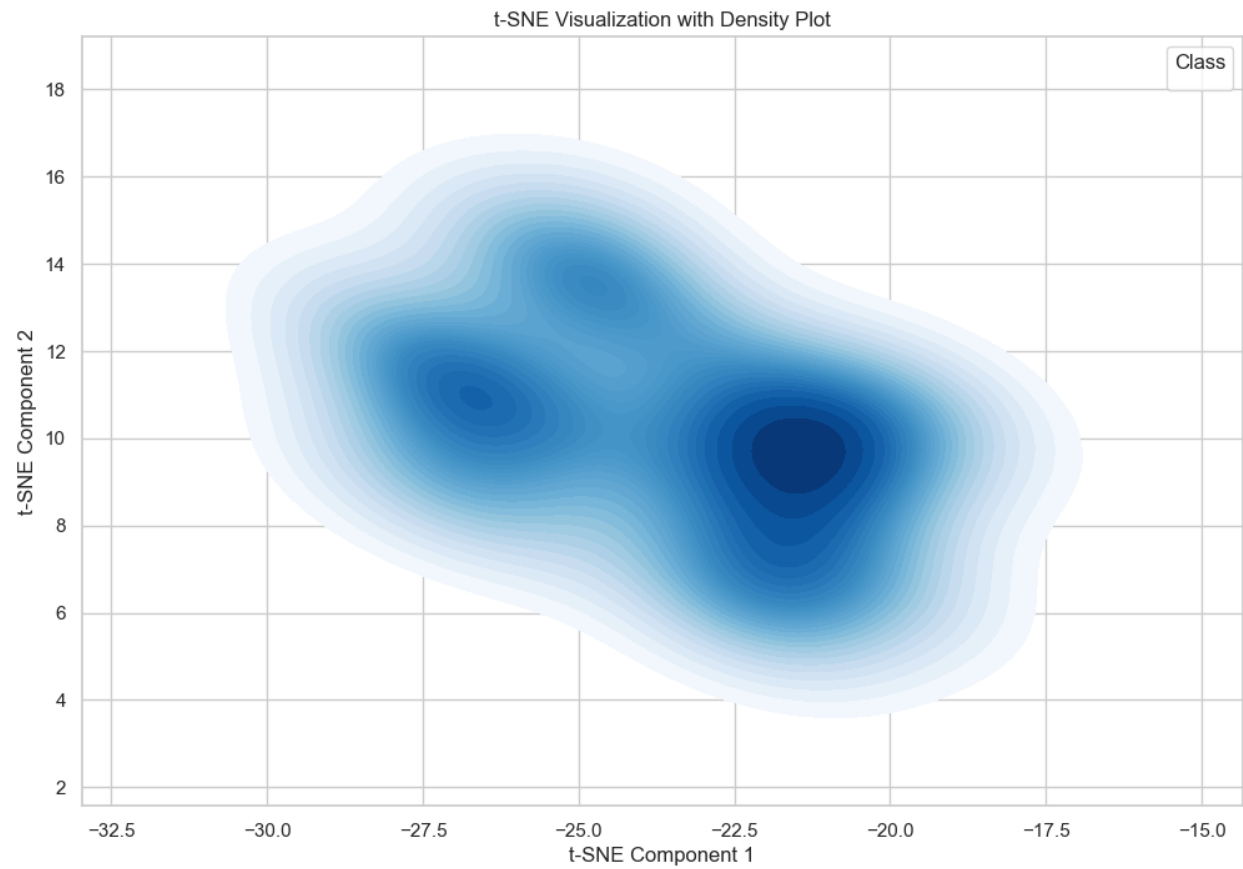


Figure 10: Density plot for t-SNE

References

- Adebamowo, C. A., et al. (2013). Epidemiology of prostate cancer in sub-Saharan Africa. *J Natl Cancer Inst Monographs*, 36(1), 27-37.
- Opoku-Agyeman, O., et al. (2017). Prostate cancer in Africa: challenges and opportunities for prevention and control. *Ann Oncol*, 28(2), 250-257.
- Parkin, D. M., et al. (2016). Global cancer statistics, 2016. *CA Cancer J Clin*, 66(5), 371-403.