The dataset that was used for the implementation of all feature selection methods (ANOVA, Chi-Square, PERMANOVA and SVM) was provided by our client and is confidential. Therefore, the csv file which includes the dataset is not included in the repository. Our repository includes the following files and folders:

Comparison_of_results: Includes the comparison results of the resulting features of all feature selection methods (ANOVA, Chi-square, PERMANOVA and SVM)

- Cluster maps
    - ANOVA
        1. ANOVA_Cluster_100_OTUs.png – cluster map created with 100 resulting OTUs of ANOVA
        2. ANOVA_Cluster_200_OTUs.png – cluster map created with 200 resulting OTUs of ANOVA
        3. ANOVA_Cluster_50_OTUs.png – cluster map created with 50 resulting OTUs of ANOVA
    - CHI2 (Chi-square)
        1. CHI2_Cluster_100_OTUs.png – cluster map created with 100 resulting OTUs of Chi-square
        2. CHI2_Cluster_200_OTUs.png – cluster map created with 200 resulting OTUs of Chi-square
        3. CHI2_Cluster_50_OTUs.png – cluster map created with 50 resulting OTUs of Chi-sqaure
    - Permanova
        1. Permanova_Cluster_100_OTUs.png – cluster map created with 100 resulting OTUs of PERMANOVA
        2. Permanova_Cluster_200_OTUs.png – cluster map created with 200 resulting OTUs of PERMANOVA
        3. Permanova_Cluster_50_OTUs.png – cluster map created with 50 resulting OTUs of PERMANOVA
    - SVM
        1. SVM_Cluster_100_OTUs.png – cluster map created with 100 resulting OTUs of SVM
        2. SVM_Cluster_200_OTUs.png – cluster map created with 200 resulting OTUs of SVM
        3. SVM_Cluster_50_OTUs.png – cluster map created with 50 resulting OTUs of SVM
    - Clustermap.py – includes the code for creating the cluster maps of the resulting features of each feature selection method by using seaborn library.
- Scatterplots
    - ANOVA
        1. ScatterPlot ANOVA 100 OTUs.png – scatterplot of 100 resulting OTUs of ANOVA
        2. ScatterPlot ANOVA 200 OTUs.png – scatterplot of 200 resulting OTUs of ANOVA
        3. ScatterPlot ANOVA 50 OTUs.png – scatterplot of 50 resulting OTUs of ANOVA

- o Chi-Square
  - 1. Chi-square 100 OTUs scatterplot.png – scatterplot of 100 resulting OTUs of Chi-square
  - 2. Chi-square 200 OTUs scatterplot.png – scatterplot of 200 resulting OTUs of Chi-square
  - 3. Chi-square 50 OTUs scatterplot.png – scatterplot of 50 resulting OTUs of Chi-square
- o PERMANOVA
  - 1. PERMANOVA Scatterplot 100 OTUs.png – scatterplot of 100 resulting OTUs of PERMANOVA
  - 2. PERMANOVA Scatterplot 200 OTUs.png – scatterplot of 200 resulting OTUs of PERMANOVA
  - 3. PERMANOVA Scatterplot 50 OTUs.png – scatterplot of 50 resulting OTUs of PERMANOVA
- o SVM
  - 1. SVM_Scatter_100.png – scatterplot of 100 resulting OTUs of SVM
  - 2. SVM_Scatter_200.png – scatterplot of 200 resulting OTUs of SVM
  - 3. SVM_Scatter_50.png – scatterplot of 50 resulting OTUs of SVM
- • Venn Diagrams
  - o 100_OTUs
    - 1. Venn_Result_100_OTUs.png – VENN diagram of the resulting 100 OTUs of all feature selection methods
    - 2. Venn_Result_OTUs_Names_100.png – table which explains the common OTUs between all feature selection methods
  - o 200_OTUs
    - 1. Venn_Result_200_OTUs.png – VENN diagram of the resulting 200 OTUs of all feature selection methods
    - 2. Venn_Result_OTUs_Names_200.png – table which explains the common OTUs between all feature selection methods
  - o 50_OTUs
    - 1. Venn_Result_50_OTUs.png – VENN diagram of the resulting 50 OTUs of all feature selection methods
    - 2. Venn_Result_OTUs_Names_50.png – table which explains the common OTUs between all feature selection methods

Feature_selection_methods_not_working – It contains all of code and other things that we tried and did not work the way we wanted it to work.

- • ANOVA – Not Working
  - o ANOVA with Statsmodels
    - 1. 1st.png – screenshot of the code tried with statsmodels library to implement ANOVA feature selection method and its results
    - 2. Explanation – Statsmodels.txt – explanation of the code and why it did not work for us

- o ANOVA
  1. Output Screenshots
     - 1st.png – screenshot of the output
     - 2nd.png – screenshot of the output
     - 3rd a.png – screenshot of the output
     - 3rd b.png – screenshot of the output
     - 3rd c.png – screenshot of the output
     - 4th.png – screenshot of the output
     - 5th.png – screenshot of the output
     - ANOVA_code.py – includes the code for ANOVA using Pingouin library that did not work for us
  2. ANOVA_Code – didn't work well.py – includes the code for ANOVA implementation using Pingouin library that did not work for us
- o Code and Output (ANOVA)- didn't work well with Pingouin
  1. Output Screenshots
     - 1st.png – includes code and results of ANOVA implementation using Pingouin library
     - 2nd.png – includes code and results of ANOVA implementation using Pingouin library
     - 3rd.png – includes code and results of ANOVA implementation using Pingouin library
     - 4th.png – includes code and results of ANOVA implementation using Pingouin library
     - Explanation – Pingouin.txt – includes explanation of why Pingouin did not work for us
- o ANOVA_using_Pingouin.py – includes code for ANOVA implementation using Pingouin library in a different way
- o ANOVA_using_Scipy.py – includes code for ANOVA implementation using SciPy python library
- Comparison_of_Results/HeatMaps
  - o Results
    1. ANOVA
       - ANOVA_100_OTUs.png – heatmap of the resulting 100 OTUs of ANOVA
       - ANOVA_200_OTUs.png – heatmap of the resulting 200 OTUs of ANOVA
       - ANOVA_50_OTUs.png – heatmap of the resulting 50 OTUs of ANOVA
    2. CHI2 (Chi-Square)
       - CHI2_100_OTUs.png – heatmap of the resulting 100 OTUs of Chi-square
       - CHI2_200_OTUs.png – heatmap of the resulting 200 OTUs of Chi-square
       - CHI2_50_OTUs.png – heatmap of the resulting 50 OTUs of Chi-square

3. Permanova
   - Permanova_100_OTUs.png – heatmap of the resulting 100 OTUs of PERMANOVA
   - Permanova_200_OTUs.png – heatmap of the resulting 200 OTUs of PERMANOVA
   - Permanova_50_OTUs.png – heatmap of the resulting 50 OTUs of PERMANOVA
4. SVM
   - 100_OTUs
     - ❖ SVM_100_OTUs.png – heatmap of the resulting 100 OTUs of SVM
     - ❖ SVM_100_OTUs_1.png – heatmap of the resulting 100 OTUs of SVM
   - 200_OTUs
     - ❖ SVM_200_OTUs.png – heatmap of the resulting 200 OTUs of SVM
   - 50_OTUs
     - ❖ SVM_50_OTUs.png – heatmap of the resulting 50 OTUs of ANOVA
     - ❖ SVM_50_OTUs_1.png – heatmap of the resulting 50 OTUs of ANOVA
- Heatmap_Visualization.py – includes code for creating the heatmaps for the resulting features of ANOVA, Chi-Square, PERMANOVA and SVM feature selection method

- Permanova
  - Distance_Matrix_Calc.py – includes code to create a distance matrix for PERMANOVA feature selection method
- SVM_Not_Working
  - SVM_not_working.py – includes code for the implementation of SVM feature selection method that did not work well with the selected dataset

Feature_selection_methods_with_sample_datasets

- ANOVA_Samples
  - o Sample 1
    1. 1st.png – screenshot of the output
    2. 2nd.png – screenshot of the output
    3. 3rd.png – screenshot of the output
    4. 4th.png – screenshot of the output
    5. 5th.png – screenshot of the output
    6. 6th.png – screenshot of the output
    7. plants_leaves – original.csv – dataset used
    8. plants_leaves.csv – transposed dataset
    9. sample1.py – includes code for the implementation of ANOVA feature selection method using pingouin, matplotlib, seaborn, pandas libraries and sample dataset
  - o Sample 2
    1. 2nd.png – screenshot of the output
    2. 3rd.png – screenshot of the output
    3. 4th.png – screenshot of the output
    4. 5th 1.png – screenshot of the output
    5. 5th 2.png – screenshot of the output
    6. Sample2.py – includes code for the implementation of ANOVA feature selection method using SciPy, matplotlib libraries and random dataset
  - o Sample_3
    1. 1st.png – screenshot of the output
    2. 2nd.png – screenshot of the output
    3. 3rd.png – screenshot of the output
    4. PlantGrowth.csv – dataset used
    5. sample3.py – includes code for the implementation of ANOVA feature selection method by using sample dataset and SciPy library

- CHI2 (Chi-square)
  - o Sample_Data_CHiSq.py – includes code for the implementation of Chi-Square feature selection method using random data and SciPy python library
- SVM_Sample
  - o Output_1.png – visuals of the results
  - o Output_2.png – visuals of the results
  - o Output_3.png – visuals of the results
  - o Output_4.png – visuals of the results
  - o Output_5.png– visuals of the results
  - o SVM_Sample.py – includes code for the implementation of SVM feature selection method by using a sample dataset
  - o breast_cancer_data.csv – sample dataset (Breast cancer data) used

Feature_selection_methods_working

- ANOVA
  - ANOVA.py – includes code for the implementation of ANOVA feature selection method
  - versions_of_libraries_used_for_ANOVA_implementation.txt – includes versions of all the libraries used for the implementation of ANOVA
- CHI
  - Chi_sq.py – includes code for the implementation of Chi-Square feature selection method
  - versions_of_libraries_used_for_Chi-square_implementation.txt – includes versions of all the libraries used for the implementation of Chi-Square
- Results
  - ANOVA
    1. output_ANOVA_100.csv – includes the list of best 100 OTUs selected by ANOVA feature selection method
    2. output_ANOVA_200.csv – includes the list of best 200 OTUs selected by ANOVA feature selection method
    3. output_ANOVA_50.csv – includes the list of best 50 OTUs selected by ANOVA feature selection method
  - CHI2
    1. Chi-OUT_50.csv – includes the list of best 50 OTUs selected by Chi-square feature selection method
    2. fnames_100.csv – includes the list of best 100 OTUs selected by Chi-square feature selection method
    3. fnames_200.csv – includes the list of best 200 OTUs selected by Chi-square feature selection method
  - Permanova
    1. Output_PERMANOVA_100_OTUs.csv - includes the list of best 100 OTUs selected by PERMANOVA feature selection method
    2. Output_PERMANOVA_200_OTUs.csv - includes the list of best 200 OTUs selected by PERMANOVA feature selection method
    3. Output_PERMANOVA_50_OTUs.csv - includes the list of best 50 OTUs selected by PERMANOVA feature selection method
  - SVM
    1. SVM_Results_100.csv – includes the list of best 100 OTUs selected by SVM feature selection method
    2. SVM_Results_200.csv – includes the list of best 200 OTUs selected by SVM feature selection method
    3. SVM_Results_50.csv – includes the list of best 50 OTUs selected by SVM feature selection method
- SVM
  - SVM.py – includes code for the implementation of SVM feature selection method
  - Requirements.txt – includes all the versions of the libraries used for the implementation of SVM feature selection method