MLBA Assignment-2

SUBHANI SHAIK (MT18117)

A. sPTD dataset

1.

Pre-processing and Feature selection:

- This dataset has two classes (Control, sPTD). The training data has 237 samples each of 29459 features after removing GA, GADel, TTD features from the original dataset.
- Analysed the dataset with different Feature selection technique like CHI2, mutual_info_classif, f1_classif, TruncatedSVD (K-SVD) and genetic algorithm. Among those f1_classif give relatively better features.
- Applied the **f1_classif** feature selection technique. By using this, selected the top 150 features which are important on and exported to MT18117_SUBHANI_SHAIK_question1**.csv** file for our machine learning model.

2.

Machine learning model:

- In this, loaded the selected features dataset and applied different machine learning models.
- I analysed and applied the :
    - Logistic regression,
    - SVM (RBF kernel)
    - Decision tree
    - MLP classifier
    - Random Forest classifier model
- For these, I tuned the parameters along with 5-fold cross-validation.
- I evaluated these models by using metrics like Accuracy, Precision, Recall, F1-score, Specificity, MCC, AUROC value.

3.

Evaluation of model:

- On these models, I evaluated them using metrics like like Accuracy, Precision, Recall, F1-score, Specificity, MCC, AUROC value.

- Most of these models giving >70% score for Accuracy, precision, recall, and f1-score.
- Some of them giving less score for specificity between 48 to 70 ranges.
- For AUROC values, they are giving relatively good results as we removed our main features.
- Exported the results to .csv file.

Note: See .csv file, PyNotebook for results and plots

4.

GA, GADel, TTD features:

- Among these features, tried the combination of these features including previous extracted features.
- These are giving better results than above(Q2,Q3) results.
- Compared these results with above results.
- Analysed and tuned the parameters according to our dataset and trained the model and taken evaluation metrics.
- Exported these results to MT18117_question4.csv file.

Note : see .csv file and PyNotebook for results and plots.

B. PRROM dataset

5.

Pre-processing and Feature selection:

- This dataset has two classes (Control, sPTD). The training data has 237 samples each of 29459 features after removing GA, GADel, TTD features from the original dataset.
- Analysed the dataset with different Feature selection technique like CHI2, mutual_info_classif, f1_classif, TruncatedSVD (K-SVD) and genetic algorithm. Among those f1_classif give relatively better features.
- Applied the **f1_classif** feature selection technique. By using this, selected the top 150 features which are important on and exported to MT18117_SUBHANI_SHAIK_question1**.csv** file for our machine learning model.
- Changed the labels like Control=1 and sPTD=0.

6.

Machine learning model:

- In this, loaded the selected features dataset and applied different machine learning models.
- I analysed and applied the :
    - Logistic regression,
    - SVM (RBF kernel)
    - Decision tree
    - MLP classifier
    - Random Forest classifier model
- For these, I tuned the parameters along with 5-fold cross-validation.
- I evaluated these models by using metrics like Accuracy, Precision, Recall, F1-score, Specificity, MCC, AUROC value.

7.

Evaluation of model:

- On these models, I evaluated them using metrics like like Accuracy, Precision, Recall, F1-score, Specificity, MCC, AUROC value.
- Most of these models giving >70% score for Accuracy, precision, recall, and f1-score.
- Some of them giving less score for specificity between 48 to 70 range.
- For AUROC values, they are giving relatively good results as we removed our main features.
- Exported the results to .csv file.

Note: See .csv file, PyNotebook for results and plots

8.

GA, GADel, TTD features:

- Among these features, tried the combination of these features including previous extracted features.
- These are giving better results than above (Q6,Q8) results.
- Compared these results with above results.

- Analysed and tuned the parameters according to our dataset and trained the model and taken evaluation metrics.
- Exported these results to MT18117_question8.csv file.

Note : see .csv file and PyNotebook for results and plots.